Computational Chemistry Methods

Applications to

Racemate Resolution and Radical Cation Chemistry

ISBN 90-9015235-0

Computational Chemistry Methods

Applications to Racemate Resolution and Radical Cation Chemistry

een wetenschappelijke proeve op het gebied van de

Natuurwetenschappen, Wiskunde en Informatica

Proefschrift

ter verkrijging van de graad van doctor
aan de Katholieke Universiteit Nijmegen,
volgens besluit van het College van Decanen
in het openbaar te verdedigen op
dinsdag 22 januari 2002,
des namiddags om 1.30 uur precies

door

Gijsbert Schaftenaar

geboren op 28 augustus 1956 te Harderwijk.

Promotores:	Prof. dr. ir. A. van der Avoird
	Prof. dr. E. Vlieg
Copromotor:	Prof. dr. R.J. Meier

Leden manuscriptcommissie:
Prof. dr. G. Vriend
Prof. dr. R.A. de Groot
Dr. ir. P.E.S. Wormer

The research reported in this thesis was financially supported by the Dutch Organization for the Advancement of Science (NWO) and DSM.

Contents

Preface

This thesis deals with the application of computational models to solve real life chemical problems. Two distinct problems are tackled. First the prediction of lattice energy differences between a pair of diastereomeric salts and secondly the elucidation of the unimolecular chemistry of the methyl carbamate radical cation. While the second problem deals with relatively small molecules (10 atoms), the first deals with much larger molecules (116 atoms). Both systems were at the limit of the size that could be handled with Quantum Mechanical methods at the time the calculations were performed. This illustrates the evolution of both Quantum Mechanical methods and computer hardware in the ten years that lie between them. Besides being different in size, these systems also differ in complexity. While the second problem deals with isolated molecules in the gas phase, the first deals with interacting molecules in the solid. In addition the latter involves a whole spectrum of intermolecular interactions, from weak (van der Waals contacts, ring-ring interactions) to strong (ionic interactions and hydrogen bridges). The total energy differences between diastereomeric salts, however, are very small (0-3 kcal/mol) compared to the energy differences found between the isomers and transition states of the methyl carbamate radical cation (up to 45 kcal/mol). Trying to reproduce such small energy differences as 0-3 kcal/mol pushes even present day computational methods to their limits.

Chapter 1

Introduction

This chapter deals with the background of racemate resolution via the formation of diastereomeric salts. In addition, a brief overview will be given of mass spectrometry techniques used for the elucidation of the unimolecular chemistry of the methyl carbamate radical cation. A review of the applied computational methods will also be given.

1.1 Introduction

1.1.1 Chirality

A chiral molecule is a molecule that is not superimposable on its mirror image. These non-superimposable mirror images are called enantiomers. Enantiomer pairs of a particular molecule have identical physical properties except for optical rotation. That is to say, when plane-polarized light is passed through a solution or crystal of one enantiomer, the resulting rotation of the light polarization plane is equal in magnitude but in opposite direction to that of its enantiomeric counterpart. For this reason, enantiomers are also called optical isomers.

The prefixes + and − or *dextro* and *laevo* (D/L), which designate the direction of the angle of rotation, are used to distinguish enantiomers. Another classification system, the R/S system, designates the absolute configuration of a stereoisomer. It has largely replaced the D/L notation and is used for molecules other than amino acids. Sometimes, the number of chiral centers (n) present in a molecule is more than one. In this case not all of the possible 2^n stereoisomers are each other's mirror image. Two stereoisomers that are not each other's mirror image are called a diastereomeric pair.

In addition to essentially identical physical properties, enantiomers also have identical chemical properties except in a chiral environment, such as biological

systems, where receptors, enzyme systems, and so forth typically have chiral properties themselves. The interaction of a drug molecule with the receptor or the enzyme is very specific due to their chemical and structural complementarity and may exhibit stereoselectivity.[1]

In fact, many active pharmaceutical drugs are chiral and were marketed up till the 1980's as racemic mixtures, i.e. as an equal mole ratio of their individual enantiomers. The problem is that one enantiomeric form of a chiral drug may be medicinally beneficial while the other enantiomeric form may be completely useless or even toxic (Table 1.1 shows some examples of the distinct biological effects of the two enantiomers in a racemate).

An example of the latter is the drug, thalidomide, which was administered to pregnant women in the 1960's (curing morning sickness in early pregnancy). One of the enantiomeric forms was found to be medicinally beneficial while the other was found to be teratogenic. The result was babies born with severe limb deformities. Since 1992 the US Food and Drugs Administration (FDA)[2,3] and the European Committee for Proprietary Medicinal Products[4] have required manufacturers to research and characterise each enantiomer in all drugs proposed to be marketed as a mixture, thus justifying that no safety-risk exists for the racemate. From that date, production of new racemates ceased to be a rational commercial option and instead became a high risk route for pharmaceutical compagnies. In addition it led to racemic-switching by pharmaceutical compagnies; extending a patent protection on a racemic drug by later patenting its single active enantiomer.

When the FDA voted in favour of single isomers, it did so because scientific advances had driven chiral technology to the point where it became realistic and routinely possible to develop them.

1.1.2 Methods for obtaining pure enantiomers

There are three options for introducing chirality into a synthesis:

- the chirality pool, where the required configuration is present in the starting materials used and is maintained throughout the remainder of any synthesis;
- asymmetric synthesis, where the single-isomer product is derived by introducing the asymmetry directly into a non-chiral material.

Table 1.1: Examples of enatiomers exhibiting different biological effects

Compound	Effect of (S)-enantiomer	Effect of (R)-enantiomer
Thalidomide[5]	teratogenic	cures morning sickness
Propanolol[6]	β -blocker	contraceptive
Limonene[7]	lemon smell	orange smell
Dopa[8]	Anti-Parkinson	toxic (granulocytopenia)
Asparagine[7]	tastes bitter	tastes sweet
Ketamine[9]	Anesthetic	responsible for side-effects (hallucination,agitation)

- resolution methods, where the precursor or material is provided as a racemic mixture and has to be separated to give the required isomer. In favourable cases the undesired isomer can be used either by turning it back into racemate which can be resolved again or by inverting its configuration so that it too provides the required isomer.

chirality pool methods

The simplest access to single isomers is their direct isolation from natural sources and in all cases the inherent chirality of nature — in plants, animals or microorganisms — has been used. An example is the anticancer drug Taxol, present in the Pacific Yew tree. Taxol's structural complexity makes total chemical synthesis impractical for drug supply. However, synthetic chemistry can help by attaching a side chain that is key to Taxol's activity onto a more abundant intermediate (Baccatin III) from the tree.[10]

Other examples are the penicillins and cholesterol lowering agent lovastatin which are derived by microbial fermentation.[11] In both cases synthetic chemistry

has also been applied to modify the fermentation products to give better drugs (such as amoxicillin[12] and simvastatin,[13] respectively, which have modified side chains). Apart from the production of complete drugs, nature also makes available to the chirality pool useful building blocks like the natural amino acids or sugars. L-Aspartic acid, for example, is used to produce the sweetener Aspartame.[14]

asymmetric synthesis from a prochiral substrate

The usual form of asymmetric synthesis takes a substrate containing no chiral elements (one that is flat) and transforms it via an asymmetric step into a chiral product. This has the advantage that potentially all the material can be realised as the required isomer directly. Such syntheses are often catalyzed, either by enzymes[15] (fermentation) or by non-enzymatic catalyst. For instance a rhodium-diphosphine catalyst is used to manufacture the anti-parkinsonian agent leveodopa.[16]

resolution of racemates

There are four ways to resolve a racemic mixture in practice: (i) *Resolution via direct crystallisation.* The racemate segregates into two morphologically different crystals, which can be separated either by hand (Pasteur[17]) or mechanically.[18] Direct crystallisation can also be achieved by seeding a supersaturated solution with an optically pure crystal of one enantiomer. Crystals containing an excess of the added enantiomer are often obtained. This procedure is called resolution by entrainment.[18] (ii) *Kinetic resolution* where the difference in reaction rate of two enantiomers with a single chiral enantiomer of another substance causes enantiomeric enrichment.[18] (iii) *Chromatographic separation.* The enantiomers form intermediate diastereomeric complexes with a chiral selector which may be present in either the stationary or the mobile phase. The difference in stability between these intermediate diastereomeric complexes results in different retention behaviour thus enabling separation.[19] (iv) *Preferential crystallization of diastereomeric salts.* A racemic acid or base is combined with a optically pure base or acid (the resolving agent), resulting in the formation of two diastereomeric salts. If these salts differ sufficiently in solubility, they can be separated by selective

crystallization of one of the salts.[18]

1.1.3 Racemate Resolution via diastereomeric salt formation

This method remains the most industrially applied method. Typically, diastereomeric salt formation has broad applicability since both enantiomers are generally available and, if used in conjunction with in-situ racemization, allows the enantiomeric yield to be significantly improved over the theoretical 50 % limit. A disadvantage of this approach was that "it was more of an art than a science" requiring a high degree of screening. Relatively recently the requirement for comprehensive screening of resolving agents for a specific substrate was reduced by the so called "Dutch resolution" technique.[20] This approach uses a mixture of structurally different resolving agents to first determine the most favored diasteromeric salt (judged by proton NMR of the crude mixture). A mixture ("family") of derivatives of the most favored structural class is used as resolving agent. For instance a mixture of mandelic, malic and dibenzoyl tartaric acids could be used as a preliminary screen (mixture). If dibenzoyl tartaric acid was found to be the most favored salt, then a mixture of functionalised dibenzoyl tartaric acids would be used as the "family". The advantages of this approach is that in more than 95 % of all cases studied (more than 300 examples by 1998) the use of the "family" approach results in higher enantiomeric excess of the substrate than a single resolving agent alone. A disadvantage, however, is that the increased amount of documentation requirements by the Federal Drug Administration limits the use of this approach for large scale manufacturing.

1.2 Rationalization of diastereomeric salt formation

The resolution by diastereomeric salt formation is based on the solubility difference in a pair of diastereomeric salts. Frank Leusen[21] proposed a model that links resolution efficiency to the lattice energy difference in a diastereomeric pair of salts. Empirically the following relation was found:

$$\Delta\Delta H_f = r_1 \ln c_p/c_n + r_2$$

Where:

- $\Delta\Delta H_f$ is the difference in heat of fusion of the diastereomeric salt-pairs,
- $\ln c_p/c_n$ is the resolution efficiency,
- c_p and c_n are the solubilities of p-salt and n-salt,
- r_1 and r_2 are constants.

The p-salt is the $(++)$ or the $(--)$ salt whereas the n-salt is the $(+-)$ or the $(-+)$ salt. Figure 1.1 shows the relation between $\Delta\Delta H_f$ and $\ln c_p/c_n$ for substituted ephedrine with substituted cyclophosphoric acid as a resolving agent (the figure was taken from the thesis of Leusen).
From the Born-Haber fusion cycle it can be deduced that the difference in lattice enthalpy ΔH_{solid} equals minus the difference in enthalpy of fusion, if we assume that the difference in liquid enthalpies is small with respect to that of the solids. Likewise, it can be deduced that the difference in lattice enthalpy ΔH_{solid} equals minus the difference in solvatation enthalpies, if we assume that the difference in enthalpy of the solutions is negligibly small:

$$-\Delta H_{solid} \approx \Delta\Delta H_f \approx \Delta\Delta H_{solv} = r_1 \ln c_p/c_n + r_2$$

This thermodynamic model was extended to include entropy effects. This gives a slightly better correlation between the entropy corrected enthalpy difference in the solid state ΔH^{corr}_{solid} and $\ln c_p/c_n$.

The resolution via diastereomeric salt formation can thus be quantified by comparing the lattice energy of the two diastereomeric salts in a pair. The lattice energy is defined as the amount of energy released when a mole of gaseous ions or

molecules are brought together from infinite separation to form a crystal. Within diastereomeric salts pairs the gas phase energy is the same. Therefore differences in lattice energies are equivalent to differences in the total energy of the solid. This opens the way to prediction of the resolution efficiency of a resolving agent by calculation of the difference in total energy of the two diastereomeric salts in a pair.

Lattice energy calculations on diastereomeric salts have been reported in the literature by two groups.

- Leusen *et al.*[21] performed molecular mechanics calculations on three pairs of

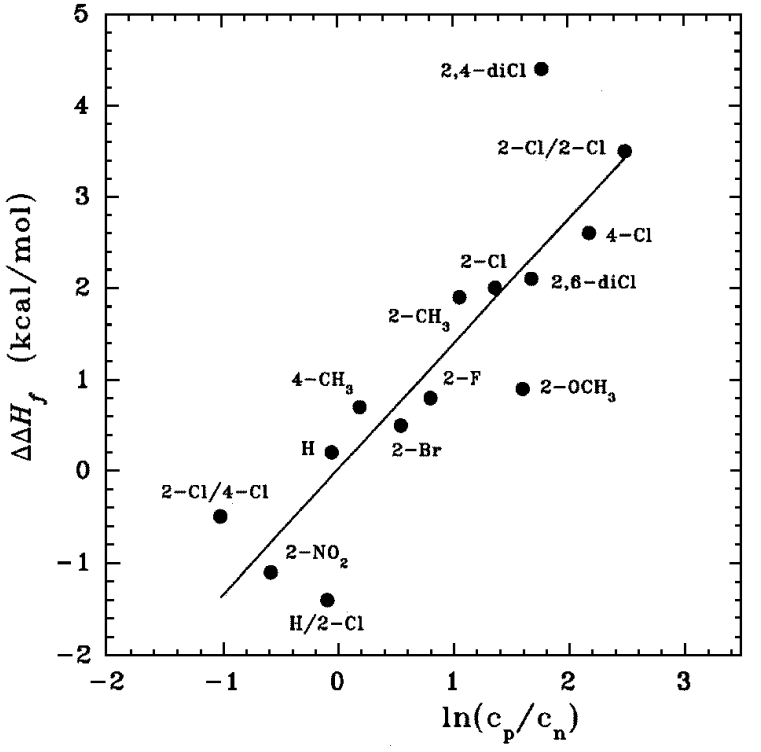


Figure 1.1: Plot of $\Delta\Delta H_f$ versus $ln(c_p/c_n)$. Labels indicate acid substituents (with unsubstituted ephedrine) or acid/base aromatic substituents.

diastereomeric salts of ephedrine with different cyclic phosphoric acids. The effect of several of the CHARMm[22] force field parameters, mainly relating to the electrostatic interaction, was investigated. These are: the effect of varying the dielectric constant, the effect of varying the cut-off distance for the electrostatic interactions and the effect of different charge schemes. The results were evaluated in terms of the ability to reproduce the X-ray structures and the energy differences of the two diastereomers in a pair. It was concluded that electrostatic potential derived charges (ESP charges) and a sufficiently large cut-off distance for the electrostatic interactions were required for a good reproduction of the X-ray structures. Still, it was not possible to obtain agreement between the experimental and calculated lattice energy differences. It was concluded that the description of the electrostatic interactions within the force field was insufficiently accurate and that the inclusion of polarization terms in the force field is essential.

- Hansen *et al.*[23] performed molecular mechanics calculations with several force fields on a wide range of compounds including diastereomeric salts. It was concluded that force fields are generally not very good at reproducing the relative stability order of polymorphs. Hansen also concluded that it is not generally possible to show that a particular charge set is superior, i.e., she showed that a given charge scheme performed well for some structures and badly for others.

From the above it is clear that an improvement in the description of the electrostatic interactions is crucial for improving the reproduction of the relative stability order of diastereomeric salts and polymorphs. The aim of our work is to find such better computational methods.
We intend to accomplish this by performing molecular mechanics calculations with improved description of the electrostatic interactions and by the use of quantum mechanical methods that can deal with periodic boundary conditions appropriate for a crystalline lattice.

We evaluate these methods on a subset of two of the diastereomeric salt pairs used by Leusen. These are the salt pairs of ephedrine with cyclic phosphoric acid and ephedrine with chlorine-substituted cyclic phosphoric acid.

1.3 Computational methods for modeling the lattice energy

This section discusses the computational methods applied; Molecular Mechanics and Quantum Mechanics methods.

1.3.1 Molecular Mechanics

The molecular mechanics method is used to calculate molecular structures and relative energies of conformers using concepts from classical mechanics. Electrons are not explicitly included in molecular mechanics, which is justified on the basis of the Born-Oppenheimer approximation stating that electronic and nuclear motions can be uncoupled from one another and considered separately. Thus, the nuclei may be viewed as moving in an average electronic potential field.

Alternatively, Molecular mechanics can be thought of as a ball-and-spring model of atoms and molecules with classical forces between them. These forces are described by potential energy functions depending on structural features such as bond lengths, bond angles, and torsional angles. The potential energy functions contain a number of parameters which are fit to reproduce experimental properties.

The total energy (E_{tot}) of a molecule or a cluster of molecules is divided into several parts, one of which is attributed to bond stretching (E_{bond}), one to bond angle bending (E_{angle}), one to torsional deformations ($E_{torsion}$), one to van der Waals interactions (E_{vdW}), and one to electrostatic interactions ($E_{Coulomb}$):

$$E^{MM} = E_{bond} + E_{angle} + E_{torsion} + E_{vdW} + E_{Coulomb} \tag{1.1}$$

A more refined force field will also include cross terms, such as stretch-bend, torsion-stretch, etc. Finally, other terms such as hydrogen bonding have been used to take into account phenomena that are not properly accounted for otherwise. The first three terms in Eq.1.1 are strictly intramolecular contributions, while the latter two are also involved in intermolecular interactions. Since the crystal packing is strongly influenced by intermolecular interactions, we will look at these in more detail.

Van der Waals potentials

The dispersion (attractive) part of the van der Waals potential is usually described by a term with an inverse sixth power of the distance, whereas the repulsive part is described by an inverse twelfth-power term (Lennard-Jones 12-6 function Eq. 1.2), or, alternatively, by an exponential function (Exponential-6 or Buckingham potential Eq. 1.3):

$$E_{vdW} = D_0\left[\left(\frac{r_0}{r}\right)^{12} - 2\left(\frac{r_0}{r}\right)^{6}\right] \tag{1.2}$$

$$E_{vdW} = A\exp(-Br) - Cr^{-6} \tag{1.3}$$

where D_0 is the well depth, r_0 is the minimum energy interaction distance, r is the interactomic distance between the interacting atoms and A, B and C are parameters.

Electrostatic potentials

The electrostatic interactions are in most force fields accounted for by a Coulombic potential energy function using partial atomic charges:

$$E_{Coulomb} = \frac{Q_i Q_j}{\epsilon\, r_{ij}} \tag{1.4}$$

where r_{ij} is interatomic distance between atoms i and j, Q_i and Q_j represent the atomic point charges, and ϵ is the MM equivalent of the dielectric constant.

Most force fields use a quick calculation of point charges based on electronegativity rules.[24,25] Several methods exist to determine point charges from a Quantum Mechanical (QM) calculation. These methods can be subdivided into a class where charges are determined by some scheme that partitions the electron density over the atoms (Mulliken,[26] Bader,[27] Hirshfeld[28]) and a class where charges are optimized to reproduce the QM electrostatic potential (ESP) by employing a least-squares fit of the model potential (on point charges based) and the QM potential. Methods in the latter class differ mainly by how and where the electrostatic potential is sampled in the surrounding molecular space.[29,30] We developed and implemented two new schemes for deriving ESP charges[31] (see chapter 3):

- Charges fit to reproduce the Distributed Multipole Analysis (DMA) derived electrostatic potential.

- ESP charges derived by sampling the electrostatic potential on a number of surfaces with constant electron density, so called isodensity surfaces.

We used the DREIDING-2.21[32] force field with a number of different charge sets to perform minimizations and lattice energy calculations on diastereomeric salts of ephedrine with cyclic phosphoric acid (see chapter 4).

The Coulomb interaction of two charge distributions can be described more accurately (than in Eq. 1.4) as a multipole expansion [33,34] in the form:

$$U_{elec} = (4\pi\epsilon_0)^{-1} \sum_{l_1 l_2 m_1 m_2} [l_1 l_2] R^{-(l_1+l_2+1)} Q_{l_1 m_1} Q_{l_2 m_2} S^{m_1,m_2}_{l_1,l_2,l_1+l_2} \qquad (1.5)$$

Where $[l_1 l_2]$ denotes the numerical factor $[(2l_1 + 2l_2 + 1)!/(2l_1)!(2l_2)!]^{1/2}$, the multipole moments Q_{lm} describe the charge distributions, and S is a function of the relative orientation of molecules.
Equation 1.5 reduces to equation 1.4 when l_1 and l_2 are zero. When one of the charge distributions is represented by only a point charge of charge +1 ($l_1 = 0, Q_{00} = 1.0$), Equation 1.5 reduces to a DMA derived electrostatic potential.

The DMAREL[35,36] package was specifically designed for crystal structure simulations using distributed multipole expansions. We used the DMAREL program to perform force field minimizations and lattice energy calculations on diastereomeric salts of ephedrine with cyclic phosphoric acid (see chapter 4).

The calculation of the electrostatic potential, ESP charges, distributed multipole moments as well as the interfacing to the DMAREL package, has been built into our program MOLDEN[37] (see chapter 2).

1.3.2 Quantum Chemical Methods

Hartree-Fock Theory and post Hartree-Fock methods

In the Hartree-Fock (HF) method the exact electronic Hamiltonian is approximated by the Fock-operator which solves an electron's wave function in the average field of all the other electrons,

$$\Big\{ -\frac{1}{2}\nabla^2 - \Big(\sum_A \frac{Z_A}{|\mathbf{r}-\mathbf{R}_A|}\Big) + \int d\mathbf{r}' \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \Big\}\ \Psi_i(\mathbf{r}) \\ - \sum_j \int d\mathbf{r}' \Big[\frac{\Psi_j(\mathbf{r}')^* \Psi_i(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \Big] \Psi_j(\mathbf{r}) = \epsilon_i \Psi_i(\mathbf{r}) \tag{1.6}$$

where the first two terms are the kinetic energy and the electron-nucleus attraction of a single electron. The third term, the *Coulomb operator*, represents the average repulsion of electron 1 by all other electrons. The last term, the *exchange operator*, arises as a direct consequence of including the Pauli principle through the use of an antisymmetrized wavefunction approximated by a single determinant wavefunction. Since the orbitals Ψ_i depend on the Coulomb and exchange operators and the Coulomb and exchange operators depend on the orbitals, the Hartree-Fock equations have to be solved iteratively until self-consistency is achieved, i.e., until the input orbitals are equal (within a certain tolerance) to the output orbitals.

In practice, the molecular orbitals (MO's) are expanded as a linear combination of atomic orbitals (AO's). The expansion coefficients are adjusted to minimize the expectation value of the total energy with the exact electronic Hamiltonian.

The primary deficiency of the Hartree-Fock theory is the inadequate treatment of the correlation between motions of electrons. While correlation of the motions of electrons with the same spin is partially taken into account by virtue of the determinantal form of the wave functions, single determinant wavefunctions do not take into account the correlation between electrons with opposite spin. As a consequence the calculated Hartree-Fock energies will be higher than the exact values. The difference between the Hartree Fock and exact energy is called the correlation energy:

$$E(\text{exact}) = E(\text{Hartree-Fock}) + E(\text{correlation}) \tag{1.7}$$

Several methods exists that can include correlation effects using the Hartree-Fock single determinant wavefunction as starting point. These post Hartree-Fock methods include configuration interaction (CI) and the Møller-Plesset perturbation series (MP2-MP4) (see Szabo and Ostlund[38]).

Density Functional Methods

The conventional ab initio methodology is based on the (approximate) solution of the many-electron Schrödinger equation. Hohenberg and Kohn[39] have shown that the total energy can directly be expressed as a functional of the electronic charge density. This concept lies at the hart of the density functional theory (DFT). Instead of working with a complex 3N-dimensional wavefunction describing the behaviour of each electron in an N-electron system, DFT allows us to work with a simple three-dimensional function, the total electronic density $\rho(\mathbf{r})$.

Kohn and Sham (KS)[40] introduced a practical approach to performing DFT calculations. In the KS approach $\rho(\mathbf{r})$ of an N-electron system (with N^α spin up and N^β spin down electrons) is expressed as the sum of the square moduli of singly occupied, orthonormal Kohn-Sham molecular orbitals Ψ_i^α and Ψ_i^β:

$$\rho(\mathbf{r}) = \rho^\alpha(\mathbf{r}) + \rho^\beta(\mathbf{r}) = \sum_i^{N^\alpha} |\Psi_i^\alpha(\mathbf{r})|^2 + \sum_i^{N^\beta} |\Psi_i^\beta(\mathbf{r})|^2 \tag{1.8}$$

The unknown energy functional, $E[\rho(\mathbf{r})]$, is partitioned into three terms:

$$E[\rho(\mathbf{r})] = U[\rho(\mathbf{r})] + T[\rho(\mathbf{r})] + E_{XC}[\rho(\mathbf{r})] \tag{1.9}$$

where $U[\rho(\mathbf{r})]$ is the classical electrostatic energy, the sum of the electron-nucleus attractions and the electron-electron Coulomb repulsions:

$$U[\rho(\mathbf{r})] = \Big(\sum_A \int -\frac{Z_A \rho(\mathbf{r})}{|\mathbf{r} - \mathbf{R}_A|}\, d\mathbf{r} \Big) + \frac{1}{2} \int\int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}\, d\mathbf{r}\; d\mathbf{r}' \tag{1.10}$$

The next term, $T[\rho(\mathbf{r})]$, is the kinetic energy of the electrons:

$$T[\rho(\mathbf{r})] = \sum_{\sigma=\alpha,\beta} \sum_i^{N^\sigma} \int \Psi_i^\sigma(\mathbf{r}) \Big[-\frac{1}{2} \nabla^2 \Psi_i^\sigma(\mathbf{r}) \Big]\, d\mathbf{r} \tag{1.11}$$

The last term, $E_{XC}[\rho(\mathbf{r})]$, is the yet to be defined exchange correlation (XC) energy functional.

Applying the condition that the energy functional is minimized by the ground state density, yields the one-electron KS equations:

$$\Big\{ -\frac{1}{2}\nabla^2 - \Big(\sum_A \frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|} \Big) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + \frac{\partial E_{XC}[\rho(\mathbf{r})]}{\partial \rho^\sigma(\mathbf{r})} \Big\} \Psi_i^\sigma(\mathbf{r}) = \epsilon_i \Psi_i^\sigma(\mathbf{r}),$$

$$\sigma = \alpha, \beta \quad (1.12)$$

Just as the Hartree-Fock equations these equations have to be solved iteratively until self-consistency is achieved, i.e., until the input densities and orbitals are equal (within a certain tolerance) to the output densities and orbitals.

Until now, we avoided dealing with the precise nature of the XC energy functional, $E_{XC}[\rho(\mathbf{r})]$. The simplest approximation to $E_{XC}[\rho(\mathbf{r})]$ is the Local Spin Density Approximation (LSDA).[41] In LSDA, $E_{XC}[\rho(\mathbf{r})]$ is approximated by:

$$E_{XC}[\rho(\mathbf{r})] = \int \rho(\mathbf{r})\ \epsilon_{XC}(\rho^{\alpha}(\mathbf{r}), \rho^{\beta}(\mathbf{r}))\ d\mathbf{r} \quad (1.13)$$

where $\epsilon_{XC}(\rho^{\alpha}(\mathbf{r}), \rho^{\beta}(\mathbf{r}))$ is the exchange and correlation energy density per particle. It is only a function of $\rho^{\alpha}(\mathbf{r})$ and $\rho^{\beta}(\mathbf{r})$ at that specific point $\mathbf{r}$ in space. The value of $\epsilon_{XC}(\rho^{\alpha}(\mathbf{r}), \rho^{\beta}(\mathbf{r}))$ has been determined for a homogeneous gas of interacting electrons as a function of the total density, by means of quantum Monte Carlo methods.[42,43] The problem with the LSDA is its systematic overestimation of binding energies[44–46] and its poor description of the hydrogen bond.[47] To solve these problems one must adopt gradient-corrected XC energy functionals. Within the Generalized Gradient Approximation (GGA), ϵ_{XC} not only depends on the density $\rho(\mathbf{r})$, but also on the gradients of the density $\nabla\rho^{\alpha}(\mathbf{r})$ and $\nabla\rho^{\beta}(\mathbf{r})$, so that $E_{XC}[\rho(\mathbf{r})]$ is now given by:

$$E_{XC}[\rho(\mathbf{r})] = \int \rho(\mathbf{r})\ \epsilon_{XC}(\rho^{\alpha}(\mathbf{r}), \rho^{\beta}(\mathbf{r}), \nabla\rho^{\alpha}(\mathbf{r}), \nabla\rho^{\beta}(\mathbf{r}))\ d\mathbf{r} \quad (1.14)$$

Several gradient-corrected exchange and correlation energy functionals exist. They are formulated to address either the exchange or the correlation component of $E_{XC}[\rho(\mathbf{r})]$. Commonly-used gradient-corrected exchange functionals are those of Perdew and Wang[48] and Becke.[49] Popular gradient-corrected correlation functionals are those of Perdew[50] and Lee, Yang and Parr (LYP).[51] Typically, a combination of one of the exchange and one of the correlation gradient-corrected XC energy functionals is used. Finally the hybrid B3LYP[52,53] functional deserves to be mentioned. The B3LYP functional is a mix of (1) the exact exchange energy, obtained in the same fashion as in a Hartree-Fock calculation, (2) the LSDA XC energy, and (3) the gradient-corrected LSDA exchange (Becke) and

correlation (LYP) energy. The weight of these three components is optimized to reproduce a set of highly accurate experimental data.[54]

QM calculations on solids

When performing QM calculations on solids the one-electron wavefunctions $\Psi_j(\mathbf{r})$ must reflect the translational symmetry of the lattice. Bloch's Theorem states that this condition is satisfied when,

$$\Psi^{\mathbf{k}}(\mathbf{r}+\mathbf{T}) = \exp(i\mathbf{k}\cdot\mathbf{T})\Psi^{\mathbf{k}}(\mathbf{r}) \tag{1.15}$$

where $\mathbf{T}$ is a direct lattice vector, $\mathbf{k}$ is a reciprocal lattice vector and $\Psi^{\mathbf{k}}$ is a Bloch function (BF). The irreducible part of the reciprocal space spanned by vectors $\mathbf{k}$ (k-space) is called the *First Brillouin Zone*. The Brillouin zone can be sampled in a finite number of $\mathbf{k}$-points to obtain good approximations to the total energy.

The one-electron orbitals in a total energy calculation are most often expanded in a set of basis functions, $\{\phi_i(\mathbf{r})\}$:

$$\Psi_j(\mathbf{r}) = \sum_i c_{i,j}\phi_i(\mathbf{r}) \tag{1.16}$$

In practical calculations one has to truncate the summation at some, preferably small, number, and it therefore becomes important to use a basis set for which the error in the representation of the orbitals is quickly converging with the number of functions. The use of a small basis set reduces the time of computation and the memory requirements. Two basic types of functions are used for the expansion in Eq.1.16: localized functions, or *atomic orbital* (AO) based BF's, and *plane waves*. For atomic orbitals, Gaussian type orbitals or numerical orbitals can be used. An example of the first is the program CRYSTAL95[55] and examples of the second are the programs DMol^3 [56] and SIESTA.[57] Atomic orbitals closely resemble the orbital behavior near the nucleus, where the valence orbitals strongly oscillate in order to stay orthogonal to the core orbitals. Many plane waves would be necessary to simulate this oscillatory behaviour. The use of pseudopotentials/pseudo-orbitals greatly reduces the number of plane waves required in the expansion, especially when ultra-soft pseudopotentials are used (see below).

The expression for the expansion of the wave functions in a plane-wave basis set is:

$$\Psi_{j,\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} c_{j,\mathbf{k}+\mathbf{G}} \ \exp[i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}] \tag{1.17}$$

The sum is over all reciprocal space vectors $\mathbf{G}$, and the wave vector $\mathbf{k}$ determines a point in the first Brillouin zone. The expansion coefficients can be shown to approach zero for large values of the kinetic energy of the plane waves, $\frac{1}{2}|\mathbf{k}+\mathbf{G}|^2$, as the energy otherwise becomes unbounded. This suggests a truncation procedure, where all plane waves below a given kinetic energy are included, and plane waves above the chosen threshold are discarded. This yields a finite number of plane waves per $\mathbf{k}$-point. VASP[58] is an example of a program that uses plane waves in combination with (ultra-soft) pseudopotentials.

Despite the fact that many more basis functions are always required to expand an orbital in plane waves, than if atom-centered basis functions were used, the use of plane waves is still advantageous for several reasons. These include the use of very efficient Fast Fourier Transform (FFT) algorithms in the iterative eigenvalue solver[59,60] as well as the conversion back and forth from real to reciprocal space. The kinetic, nuclear attraction and Coulomb repulsion contributions to the total energy are calculated in reciprocal space,[61] while the exchange-correlation contribution is calculated in real space.

The number of Fourier grid points necessary is determined by the oscillations in the density. The density requires a grid with twice the linear size of the grid required by the wave functions, because the squaring of the wave functions to obtain the density doubles the frequencies of the Fourier components.

To save time on computations one can often benefit from doing all the calculations related to the wave functions on a coarse Fourier grid, and transport the wave functions to a denser grid in reciprocal space by Fourier interpolation before doing the few calculations related to the electron density.[62] This double-grid technique becomes especially important for larger calculations in the ultrasoft scheme, because the ultrasoft wave functions are much softer than the augmentation charge density.

Pseudopotentials

One way of avoiding much of the computational expense associated with the all-electron approach is by describing the core region in terms of fixed atomic orbitals that perturb (but are not necessarily perturbed by) the valence region. This is known as the frozen core approximation. The same goal can be achieved through the use of pseudopotentials.

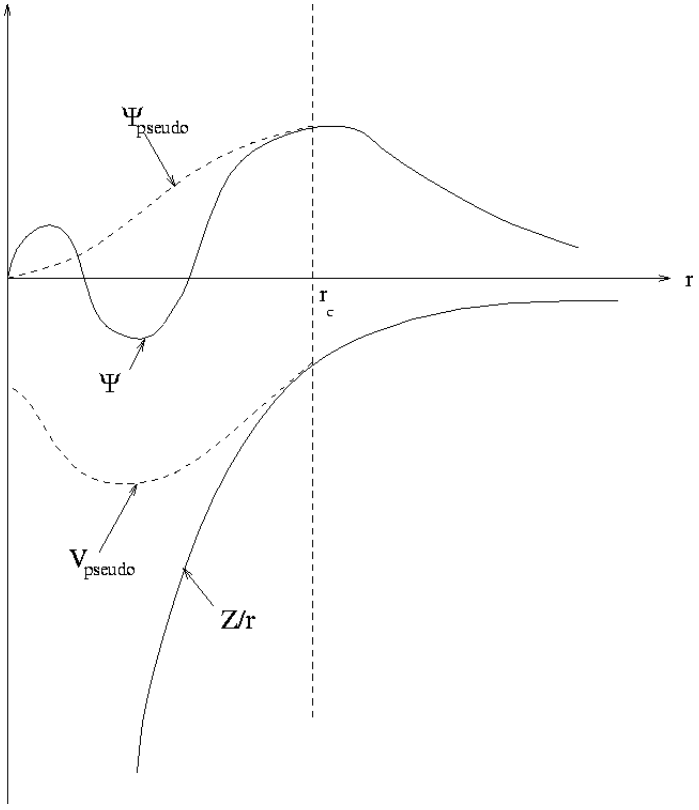


Figure 1.2: An illustration of the actual valence wavefunction Ψ and electronic potential (solid lines) plotted against distance, r, from the atomic nucleus. The corresponding pseudo-wavefunction Ψ_{pseudo} and potential is plotted (dashed lines). Outside a given radius, r_c , the actual and pseudo wavefunctions are identical.

In the norm-conserving pseudopotential approach,[63] pseudo-wavefunctions are constructed which are equal to the actual valence wavefunctions beyond some defined core radius r_c, but different inside r_c, while retaining an identical norm, i.e., net charge density within the sphere of radius r_c) (See figure 1.2).

Vanderbilt (1990)[64] removed the norm-conservation criterion of the pseudopotential, so that the pseudo-wavefunctions inside r_c could be made as soft as possible (Ultrasoft). His approach greatly reduces the number of plane waves required in the expansion of the wavefunction. Since the pseudo-wavefunctions are no longer normalized one has to correct for the missing charge density by adding so-called augmentation charge density.

1.4 Summary of used mass spectrometry techniques

Chapter 5 of this thesis deals with the unimolecular chemistry of the methyl carbamate radical cation studied by mass spectrometry techniques and ab initio molecular orbital calculations. This was a joint project between the Theoretical Chemistry and Mass Spectrometry groups of the University of Utrecht. Below is a summary of the applied mass spectrometry techniques.

In the ionization chamber of a mass spectrometer the sample M_1 is ionized by electrons. Ejection of an electron from the sample molecule leads to a radical cation:

$$M_1 + e \rightarrow M_1^+ + 2e$$

The minimum energy required for ionization is the Ionization Energy (IE). From the $IE(M_1)$ and the heat of formation[a] ΔH_f^0 of the sample M_1, the heat of formation of the ion $\Delta H_f^0(M_1^+)$ can be derived:

$$IE(M_1) = \Delta H_f^0(M_1^+) - \Delta H_f^0(M_1)$$

The molecular ion M_1^+ can also fragment into a daughter ion M_2^+ and a neutral

[a] The heat of formation is defined as the enthalpy change, ΔH_f^0, for the formation of 1 mol of a substance in the standard state from the most stable forms of its constituent elements in their standard states.

M_3:

$M_1^+ \rightarrow M_2^+ + M_3$

The energy at which the daughter ion M_2^+ first appears, is called the Appearance Energy (AE). From the $AE(M_2^+)$, the heat of formation of the ion M_2^+ can be derived via the following relation (assuming the other ΔH_f^0 's are known):

$AE = \Delta H_f^0(M_2^+) + \Delta H_f^0(M_3) - \Delta H_f^0(M_1)$

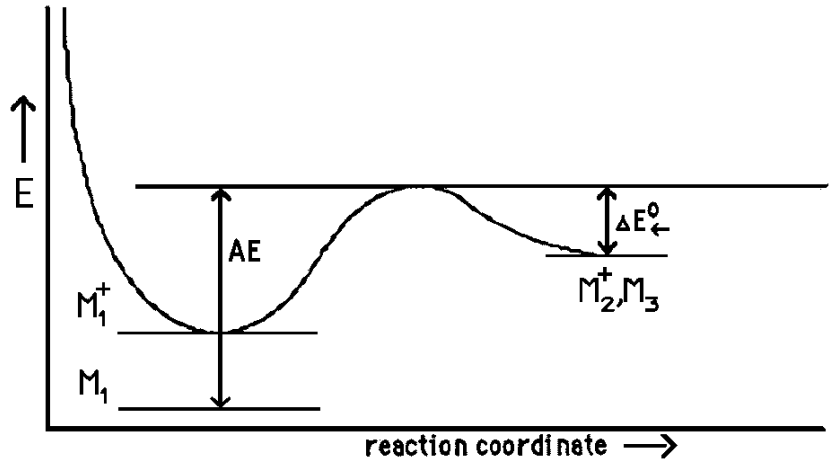


Figure 1.3: Energy diagram for the fragmentation of an ion M_1^+ into the ion M_2^+ and neutral M_3. Arrows indicate the Appearance Energy (AE) and the reverse activation energy $\Delta E_{\leftarrow}^0$

The above only holds if no reverse reaction energy barrier is involved and the heat of formation of M_1^+ lies below the heat of formation of the products (see figure 1.3). The presence of an energy barrier results in a higher value for the AE, which then yields valuable information about the barrier.[65] After passage of the barrier at least the reverse activation energy is transferred to the fragmentation product in the form of internal and kinetic energy.[66] This kinetic energy release

$\langle T \rangle$ is observable and can be derived from the broadening of the metastable ion peak. Metastable ions are ions to which relatively little energy is transferred by the ionizing electron. The lifetime of these ions is such that they can undergo rearrangements but fragment somewhere inside the mass spectrometer.

Stable ions can be forced to fragment by *Collision Activation* (CA) with a target gas. Information about the identity of an ion can also be retrieved by investigating the stability of its neutral counterpart with *Neutralization-Reionization Mass Spectrometry* (NRMS).[67] In this experiment the ion is neutralized and subsequently reionized again. Finally the neutrals released during ion fragmentation can be studied by *Collision Induced Dissociative Ionization* Mass Spectra[68] (CIDI).

1.5 Outline of this Thesis

Chapter 2 deals with the Molden package, our visualization program of molecular and electronic structure. This program has become central to this research. It is used to calculate a number of different types of charges, the electrostatic potential, Distributed Multipoles, Z-matrix constructs to impose translational and screw type of symmetry, for interfacing with DMAREL and VASP, and for manipulation of crystal structures.

Chapter 3 deals with the development of two new point-charge models:

- Charges fit to reproduce the Distributed Multipole derived electrostatic potential.
- ESP charges derived by sampling the electrostatic potential on a number of surfaces with constant electron density.

Chapter 4 deals with the Quantum Mechanical and Force Field calculations on Diastereomeric Salts.

Chapter 5 deals with the unimolecular chemistry of the methyl carbamate radical cation, investigated by a combination of mass spectrometry techniques and ab initio molecular orbital calculations.

References and Notes

[1] U. Hacksell, *A Textbook of Drug Design and Development*, ch. Structural and physicochemical fctors in drug action. New York: Harwood Academic Publishers, 1991.

[2] Chirality, 4 (1992) 338-340.

[3] M. Gross, *Chemical Analysis Vol. 142: The Impact of Stereochemistry on Drug Development and Use.*, ch. Chapter 20: Enantioselective analysis and the regulation of chiral drugs. New York: John Wiley rm & Sons, Inc., 1997.

[4] CPMP, 1994, Note for Guidance. Investigation of Chiral Active Substances. Brussels: Commision of the European Union, III/3501/91-EN Final.

[5] De Camp, W.H., Chirality, 1 (1989) 2.

[6] Vermeulen, A.M., Belpaire, F.M., Moerman, E., de Smet, F., Bogaert, M.G., Chirality, 4 (1992) 73.

[7] Enders, D., Hoffmann, R.W., Chemie in unzerer Zeit, 6 (1985) 177.

[8] Cotzias, G.C., Papavasiliou, P.S., Gallene, R., N. Engl. J. Med., 280 (1969) 337-345.

[9] White, P.F., Ham, J., Way, W.L., Trevor, A.J., Anesthesiology, 52 (1980) 231-239.

[10] Pirkle, W.H., Pochapsky, T.C., Chem. Rev., 89 (1998) 347-362.

[11] Okamoto, Y., Kaida, Y., J. Chromatogr., 666 (1994) 403-419.

[12] Nathwani, D., Wood, M.J., Drugs, 45 (1993) 866-894.

[13] Hoffman, W.F., Alberts, A.W., Anderson, P.S., Chen, J.S., Smith, R.L., Willard, A.K., J. Med. Chem., 29 (1986) 849-852.

[14] Tou, J.S., Vineyard, B.D., J. Org. Chem., 50 (1985) 4982-4984.

[15] Crosby, J., Tetrahedron, 47 (1991) 4789-4846.

[16] Knowles, W.S., Sabacky, M.J., Vineyard, B.D., Weinkauff, D.J., J. Am. Chem. Soc., 97 (1975) 2567-2568.

[17] Pasteur, L., Ann. Chim. et Phys., 24 (1848) 442.

[18] Jacques, J., Collet, A., Wilen, S.H., *Enantiomers, Racemates and Resolutions.* Malabar, Florida: Krieger Publishing Company, 1994.

[19] Taylor, D.R., Maher, K., J. Chromatographic Science, 30 (1992) 67-85.

[20] Vries, T., Wynberg, H., van Echten, E., Koek, J., ten Hoeve, W., Kellog, R.M., Broxterman, Q.B., Minnaard, A., Kaptein, B., van der Sluis, S., Hulshof, L., Kooistra, J., Angew. Chem. Int. Ed., 89 (1998) 37.

[21] F. Leusen, *Rationalization of racemate resolution: a molecular modelling study.* Nijmegen: Ph.D. thesis, 1993.

[22] Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. Karplus, M., J. Comput. Chem., 4 (1983) 187-217.

[23] L. Hansen, *Structural Investigations of Diastereomeric Salts.* The Royal Danish School of Pharmacy: Ph.D. thesis, 1997.

[24] Del Re, G., J. Chem. Soc. London, 1958 (1958) 4031.

[25] Gasteiger, J., Marsili, M., Tetrahedron, 36 (1980) 3219.

[26] Mulliken, R.S., J. Chem. Phys., 23 (1955) 1833.

[27] R. Bader, *Atoms in Molecules - A Quantum Theory.* Oxford: Oxford University Press, 1990.

[28] Hirshfeld, F.L., Theor. Chim. Acta, 49 (1977) 129.

[29] Breneman, C.M., Wiberg, K.B., J. Comp. Chem., 11 (1990) 361.

[30] Besler, B.H., Merz, K.M., Kollman, P.A., J. Comp. Chem., 11 (1990) 431.

[31] Schaftenaar, G., Noordik, J.H., J. Comp.-Aided Molecular Design, 14 (2000) 233.

[32] Mayo, S.L., Olafson, B.D., Goddard III, A., J. Phys. Chem., 94 (1990) 8897.

[33] Stone, A.J., Chem. Phys. Lett., 83 (1981) 233.

[34] Price, S.L., Stone, A.J., Alderton, M., Molec. Phys., 52 (1984) 987.

[35] Coombes, D.S., Price, S.L., Willock, D.J., Leslie, M., J. Chem. Phys., 100 (1996) 7352.

[36] Willock, D.J., Price, S.L., Leslie, M., Catlow, C.R.A., J. Comp. Chem., 16 (1995) 628.

[37] Schaftenaar, G., Noordik, J.H., J. Comp.-Aided Molecular Design, 14 (2000) 123.

[38] A. Szabo and N. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory.* New York: McMillan, 1982.

[39] Hohenberg, P., Kohn, W., Phys. Rev. B, 136 (1964) 864.

[40] Kohn, W., Sham, L.J., Phys. Rev. A, 140 (1965) 1133.

[41] Vosko, S.H., Wilk, L., Nusair, M., Can. J. Phys., 58 (1980) 1200.

[42] Ceperley, D.M., Phys. Rev. B, 18 (1978) 3126.

[43] Ceperley, D.M., Alder, B.J., Phys. Rev. Lett., 45 (1980) 566.

[44] Johnson, B.G., Gill, P.M.W., Pople, J.A., J. Chem. Phys., 98 (1993) 5612.

[45] Becke, A.D., J. Chem. Phys., 96 (1992) 2155.

[46] Becke, A.D., J. Chem. Phys., 97 (1992) 9173.

[47] Sim, F., St-Amant, A., Papai, I., Salahub, D.R., J. Am. Chem. Soc., 114 (1992) 4391.

[48] Perdew, J.P., Wang, Y., Phys. Rev. B, 33 (1986) 8800.

[49] Becke, A.D., Phys. Rev. A, 38 (1988) 3098.

[50] Perdew, J.P., Phys. Rev. B, 33 (1986) 8822.

[51] Lee, C., Wang, W., Parr, R.G., Phys. Rev. B, 37 (1988) 785.

[52] Becke, A.D., J. Chem. Phys., 98 (1993) 1372.

[53] Becke, A.D., J. Chem. Phys., 98 (1993) 5648.

[54] Becke, A.D., J. Chem. Phys., 107 (1997) 8554.

[55] Dovesi, R., Saunders, V.R., Roetti, C. Causà, M., Harrison, N.M., Orlando, R., Apra, E., Crystal-Electronic Structure Structure of Periodic Systems, User Manual. http://gserv1.dl.ac.uk/TCSC/Software/CRYSTAL/, 1996.

[56] Delley, B., J. Chem. Phys., 92 (1990) 508.

[57] Sanchez-Portal, D., Ordejon, P., Artacho, E., Soler, J.M., Int. J. of Quant. Chem., 65 (1997) 453.

[58] Kresse, G., Hafner, J., Phys. Rev. B, 47 (1993) 558.

[59] Car, R., Parrinello, M., Phys. Rev. Lett., 55 (1985) 2471-2474.

[60] Kresse, G., Furthmüller, J., Comput. Mat. Sci., 6 (1996) 15.

[61] Ihm, J., Zunger, A., Cohen, M.L., J. Phys. Chem, 12 (1979) 4409-4422.

[62] Laasonen, K., Pasquarello, A., Car, R., Lee, C., Vanderbilt, D., Phys. Rev. B, 47 (1993) 10142.

[63] Hamann, D.R., Schlüter, M., Chiang, C., Phys. Rev. Lett., 43 (1979) 1494.

[64] Vanderbilt, D., Phys. Rev. B, 41 (1990) 7892.

[65] Holmes, J.L., Mommers, A.A., de Koster, C., Heerma, W., Terlouw, J.K., Chem. Phys. Lett., 115 (1985) 437.

[66] Beynon, J.H., Gilbert, R., *Application of Transition State Theory to Unimolecular Reactions.* Chichester: Wiley Interscience, 1984.

[67] Terlouw, J.K., Schwarz, H., Angew. Chem. Int. Ed. Engl., 26 (1987) 805.

[68] Burgers, P.C., Holmes, J.L., Mommers, A.A., Terlouw, J.K., Chem. Phys. Lett., 102 (1983) 1.

Chapter 2

Molden: a pre- and post- processing program for molecular and electronic structures

2.1 Summary

Molden is a software package for pre- and postprocessing of computational chemistry program data. Interfacing to the ab initio programs GAMESS-US/UK and GAUSSIAN and to the Semi-Empirical package MOPAC is provided. The emphasis is on computation and visualization of electronic and molecular properties but, e.g., reaction pathways can be simulated as well. Some molecular properties of interest are processed directly from the output of the computational chemistry programs, others are calculated in MOLDEN before display. The package features different options to display MOLecular electronic DENsity, each foccusing on a different structural aspect: molecular orbitals, electron density, molecular minus atomic density and the Laplacian of the electron density. To display difference density, either the spherically averaged atomic density or the oriented ground state atomic density can be used for a number of standard basis sets. The quantum mechanical electrostatic potential or a distributed multipole expansion derived electrostatic potential can be calculated and atomic charges can be fitted to these potentials calculated on Connolly surface(s). Reaction pathways and molecular vibrations can be visualized. Input structures can be generated with a Z-matrix editor. A variety of graphics languages is supported: XWindows, postscript, VRML and Povray format.

2.2 Introduction

Quantum chemistry programs like **Gaussian**[1] and **Gamess-US/UK**[2,3] and the semiempirical program **Mopac**,[4] have become widely accepted as valuable tools in fields as diverse as drug design, synthesis planning and material science. The continuous speedup of computer hardware, has made calculations on systems of interest tractable to the bench chemist. This bench chemist, however, is often discouraged from using these computational tools because of the use of unfamiliar concepts or complicated user interfaces. Some vendors of computational chemistry programs are simplifing the use of their product by the development of a customized graphical interface. For example, GaussView,[5] SYBYL from Tripos[6] or CERIUS from MSI[7] provide interfaces to one or more computational chemistry packages. High quality computer graphics is playing an increasingly important role in computational chemistry. Visualization of computational results is often the most important road to interpretation.

The program Molden adds the power of computer graphics to the interpretation of the calculations. Other visualization programs feature sophisticated routines to display structural details, but Molden is quite unique in its coverage of features of the electronic structure. Molden has been designed as a general purpose tool to overcome the barriers which might hamper the use of computational chemistry techniques and programs. It facilitates the access to these programs via interactive preparation of program input.

Molden does not only visualize results produces by other programs, but it also can calculate several interesting quantities which those programs do not provide or require multiple invocations of those programs. From now on we will refer to first form of visualization as "direct visualization" and the second form as "indirect visualization". Prerequisites for indirect visualization such as, the atomic coordinates, the basis set, the molecular orbital coefficients and the number of electrons, are read from the program outputs.
Quantities calculated by Molden are:

- the electrostatic potential (ESP),
- ESP charges (a fit of point charges to reproduce the electrostatic potential),
- the Distributed Multipole Analysis (DMA),

- the orbitals, the molecular density or the difference density,
- the Laplacian of the electron density.

Other quantities visualized by Molden, directly available from the output of computational chemistry programs, include the intermediate and final results of a geometry optimisation or saddle point location, together with related information such as energies, forces and convergence data, the self consistent field convergence and normal modes.

Although mainly designed as a pre-/post- processing tool for the computational chemistry programs mentioned before and with a focus on the electronic structure,[8] Molden also provides more general applications to the user. Current options include:

- visualization of protein structures as available from the Brookhaven Protein Data Bank.[9]
- visualization of information on protein secondary structure elements such as alpha helices, beta-sheets or random coil. When such information is not available from the PDB entry, it is generated according to the method described in an early version of the *VADAR* program.[10]
- visualization of crystal structures from the Cambridge Structural Database[11] as specified in the *FDAT file format*,[12] including the generation of symmetry equivalent positions and a unit cell and the display of multiple unit cells.

Supported file formats also include, the XMOL format[13] and the Chemx format.[14]

User documentation and details of the graphics and visualization aspects of the program are available as a Web document:

URL: http://www.caos.kun.nl/~schaft/molden/molden.html

The appreciation of the program by both computational and bench chemists, may be illustrated by more than 1300 installations worldwide.

2.3 Methods

2.3.1 Molden as a preprocessor; the Z-matrix Editor

Preparation of input for computational chemistry programs often is a problem for even the "computer literate" bench chemist. The structure of an input file can be quite complex due to the large number of supported options and the sometimes complicated rules for parameter specification. Either cartesian coordinates or a Z-matrix can be used for the specification of the molecular geometry. In the Z-matrix approach, atom positions are defined with respect to previously defined atoms by means of internal coordinates such as bond distances, bond angles and dihedral angles. For small molecules a Z-matrix can often be constructed by "hand", but for larger molecules this quickly becomes tedious and complex.

Molden's Z-matrix Editor presents the user complete control over the molecular geometry. Modifications in internal variables are immediately reflected in the displayed structure and modifications in the structure are immediately incorporated in the current Z-matrix. New structures can be built, adding one atom at a time or by using internally or externally stored fragments. Internal coordinates can be specified to be variable, constant or "linked" to other internal coordinates. In the latter case two or more internal coordinates are described with one variable. This is particulary useful in the construction of Z-matrices that reflect the molecular symmetry. Geometries constructed with the Z-matrix editor can be saved as a Z-matrix or using cartesian coordinates. Figure 2.1 shows an example display with both a molecule display and the corresponding Z-matrix "window".

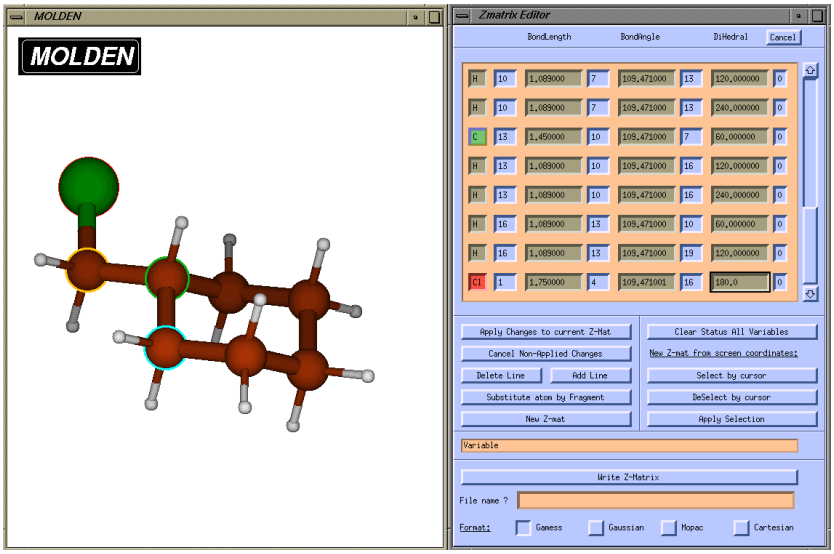


Figure 2.1: Molden's Z-matrix Editor option. The left screen window displays the "editable" molecule; the right screen window the corresponding Z-matrix and the edit-control buttons.

2.3.2 Molden as a postprocessor

Molden processes numerical information produced by computational chemistry programs. This processing can result in direct visualization, if all necessary information is already available. If such is not the case, additional visualization information will be calculated in an intermediate computational step.

Visualization of reaction paths and normal modes

The use of Molden for the animation of reaction paths and normal modes of vibration of molecules, are examples of direct visualization.

During a reaction the molecular geometry generally will change. These geometrical or conformational changes are correlated with changes in the total energy. Molden can visualize reaction paths from Internal Reaction Coordinate

(IRC) calculations, optimization and/or saddle runs. Molden complements such an animation with a plot of the energy versus geometry point.

Information on conformational flexibility resulting from thermal motion is present in the output of computational chemistry programs in the form of normal modes and their frequencies. Molden displays normal modes as a series of geometries. The starting geometry is gradually distorted by scaling the atomic displacement vectors that make up a normal mode. Molden allows the animation to be saved in the form of a VRML scene or a series of GIF files. The latter can be converted with external programs to commonly used animation file formats, such as GIF89a[15] also known as gifanim.

Figure 2.2 shows the racemization of hexahelicene as an example of direct visualization.

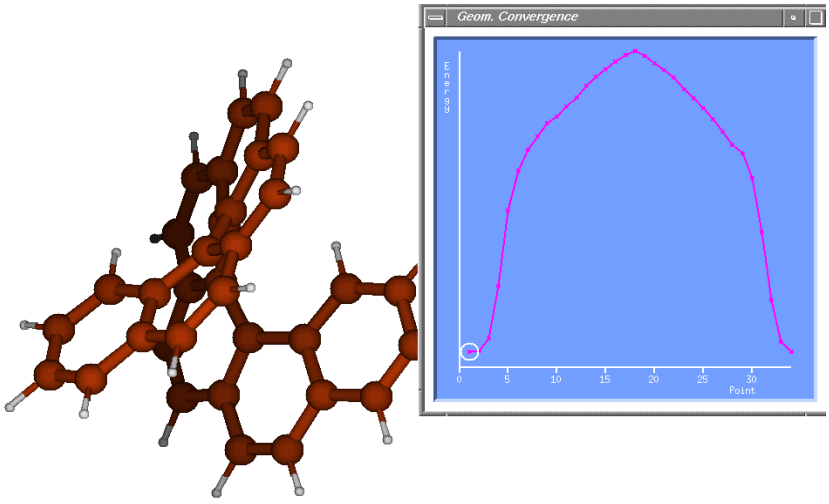


Figure 2.2: Racemization of hexahelicene. The left hand window of the screen shows the animation. In the right hand window an energy plot is displayed simultaneously. By selecting a position on the energy plot, the corresponding molecular geometry will be displayed.

Some of Molden's options for the display of electronic structural information require information not directly available from the computational chemistry program output. In particular the handling of the electron density and charge distribution are far from trivial and require closer attention. Generally one or more additional computational steps are required to generate the required information.

Visualization of electron densities

Difference density

Plots of the difference or deformation density are often used [16–18] to visualize the effects of chemical bonding. In these plots of molecular density minus spherical atomic density, the accumulation and depletion of electron density during bond formation is shown. A classic example is the formation of banana shaped bonds in cyclopropane, as shown in Figure 2.3.

For some molecules it has been shown[18,19] that the standard deformation density is small or even negative for certain covalent bonds, indicating that the use of spherically symmetric atomic ground state densities is not always the best atomic reference. Although atomic ground states are spherically symmetric, some degenerate atomic ground states are actually linear combinations of nonspherical oriented components. This degeneration can be lifted even at large interatomic distances, where the very weak long-range influence of other atoms has an orienting effect on the free atom with a degenerate ground state. In such cases it is better to use the oriented nonspherical components of the spherically symmetric ground states .[19,20,18,21,22] Molden incorporates two algorithms for the construction of the difference density maps. The first uses spherically symmetric atomic densities and the second uses oriented non spherically symmetric atomic densities. The first algorithm is described below:

The total molecular electron density σ is defined as the sum over the densities of the occupied Molecular Orbitals (**MO**'s)[23]:

$$\sigma = \sum_i nocc_i \; MO_i^2 = \sum_i \sum_r \sum_s \; nocc_i \; c_{ir} c_{is} AO_r AO_s = \sum_r \sum_s P_{rs} AO_r AO_s \quad (2.1)$$

where each MO_i is a linear combination of Atomic Orbitals (AO_r) with coefficients c_{ir}. Each MO_i is occupied with $nocc_i$ electrons. P_{rs} are the elements of

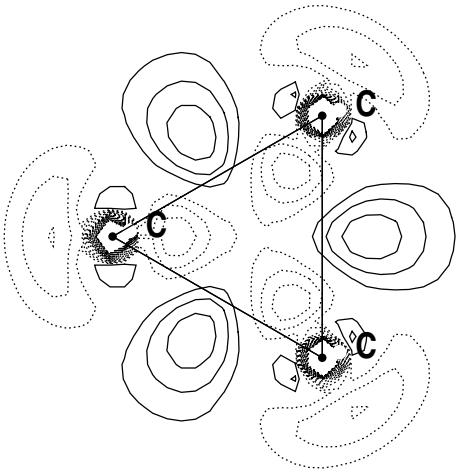


Figure 2.3: Cyclopropane sto3g basisset Molecular density minus Atomic density. The "standard" approach displays the correct density distribution with "banana bonds"; an increase of electron density positioned outside the C-C bond axes (solid contours) and a depletion of electron density in the inner ring around the carbon atoms (dashed contours).

the 1-electron density matrix:

$$P_{rs} = \sum_i nocc_i \; c_{ir} c_{is} \tag{2.2}$$

For the H_2 molecule, where the simplest wavefunction consists of a doubly occupied MO, consisting of an $1s$ orbital on atoms **a** and **b** this leads to:

$$MO = \{c_1 1s_a + c_2 1s_b\} \tag{2.3}$$

and the electron density is given by:

$$\sigma = 2\left\{c_1^2 1s_a^2 + 2c_1 c_2 1s_a 1s_b + c_2^2 1s_b^2\right\} \tag{2.4}$$

or in matrix form:

$$\begin{pmatrix} 1s_a & 1s_b \end{pmatrix} \begin{pmatrix} 2c_1^2 & 2c_1c_2 \\ 2c_2c_1 & 2c_2^2 \end{pmatrix} \begin{pmatrix} 1s_a \\ 1s_b \end{pmatrix} \tag{2.5}$$

where the matrix in the center is the 1-electron density matrix P. P can be thought of as composed of a set of atomic matrices, i.e. matrices refering to the same AO products as those available to the free atoms,

$$\begin{pmatrix} 2c_1^2 & 0 \\ 0 & 2c_2^2 \end{pmatrix} \tag{2.6}$$

and the overlap density matrix, with matrix elements refering to products of AO's centered on different atoms.

$$\begin{pmatrix} 0 & 2c_1c_2 \\ 2c_2c_1 & 0 \end{pmatrix} \tag{2.7}$$

We now define the difference or deformation density as the molecular density minus the spherical atomic density. In the case of the H_2 molecule this becomes:

$$\begin{pmatrix} 2c_1^2 - 1 & 2c_1c_2 \\ 2c_2c_1 & 2c_2^2 - 1 \end{pmatrix} \tag{2.8}$$

If an atomic groundstate is not spherical symmetric, a difference density plot will show a misleading decrease in electron density along some bonds. E.g., in the case of H_2O_2 (Figure 2.4), accumulation of charge density is absent along the O-O axis (the x-axis in the molecular frame). In this example oxygen has a ^3P ground state. One of the not spherical symmetric components is $p_x^2p_y^1p_z^1$. Here the electron density has a maximum along the x-axis. Subtracting a spherically symmetric atom $p_x^{4/3}p_y^{4/3}p_z^{4/3}$ can in such a case result in subtracting too much density in the y and z directions and too little in the x direction. In general, if an atom in a molecule has retained much of its ground state character, it will have a preferred orientation. The effect is most pronounced for the atoms O,F,S and Cl, whereas for example C usually has retained most of its spherically symmetric character.

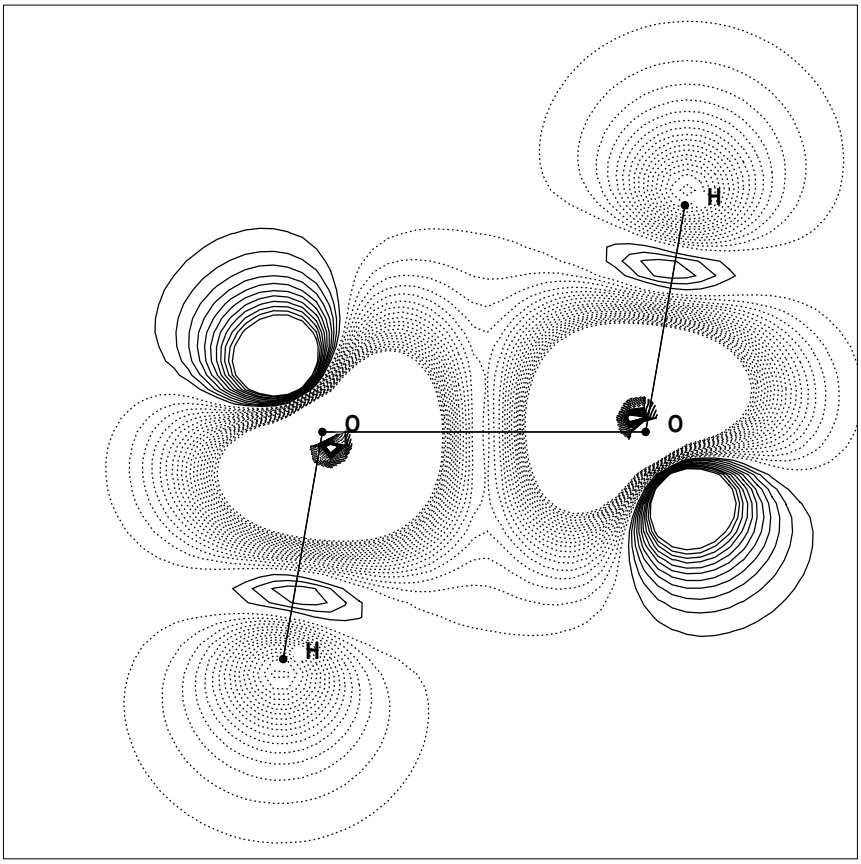


Figure 2.4: H_2O_2 sto3g molecular density minus spherical atomic density (standard approach). Misleading density plot as a result of neglecting the directional character of atomic density. Accumulation of charge density (solid contours) is absent along the O-O axis.

The algorithm in molden that uses oriented non spherically symmetric atomic densities, modifies the atomic density D_{atom} of O,F,S and Cl atoms such that the sum of $(D_{mol}(i,j) - D_{atom}(i,j))^2$ becomes minimal, with D_{mol} being the atomic part of the molecular density matrix and i and j running over the p_x, p_y, p_z atomic orbitals. Application of this approach to the H_2O_2 molecule results in the plot shown in Figure 2.5 with the expected increase in electron density along the O-O axis.

The electron density difference function to be minimized becomes in matrix form:

$$\begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{pmatrix}^{Mol} - \begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{pmatrix}^{Atom} * \begin{pmatrix} R_{xx} & R_{xy} & R_{xz} \\ R_{yx} & R_{yy} & R_{yz} \\ R_{zx} & R_{zy} & R_{zz} \end{pmatrix} \quad (2.9)$$

Where matrix R is a rotation matrix defined by three euler angles. The sum of the squares of this difference matrix $\sum_{i=x,y,z} \sum_{j=x,y,z} D_{ij}^2$ is the quantity that is actually minimized.

The atomic density matrices are stored internally for a number of frequently used basis-sets and elements.

Molecular orbitals

The second Molden option related to orbital display is the visualization of orbitals such as frontier molecular orbitals as used in Frontier Molecular Orbital (FMO) theory calculations .[24]

Figure 2.6 shows the HOMO in the Diels-Alder cycloaddition transition state of butadiene with ethylene, which FMO theory predicts to be a HOMO-LUMO reactant combination.

The Laplacian of the electron density

A third Molden option for the display of orbital related electron density information is the Laplacian of the electron density. Bader showed[25–28] how this quantity

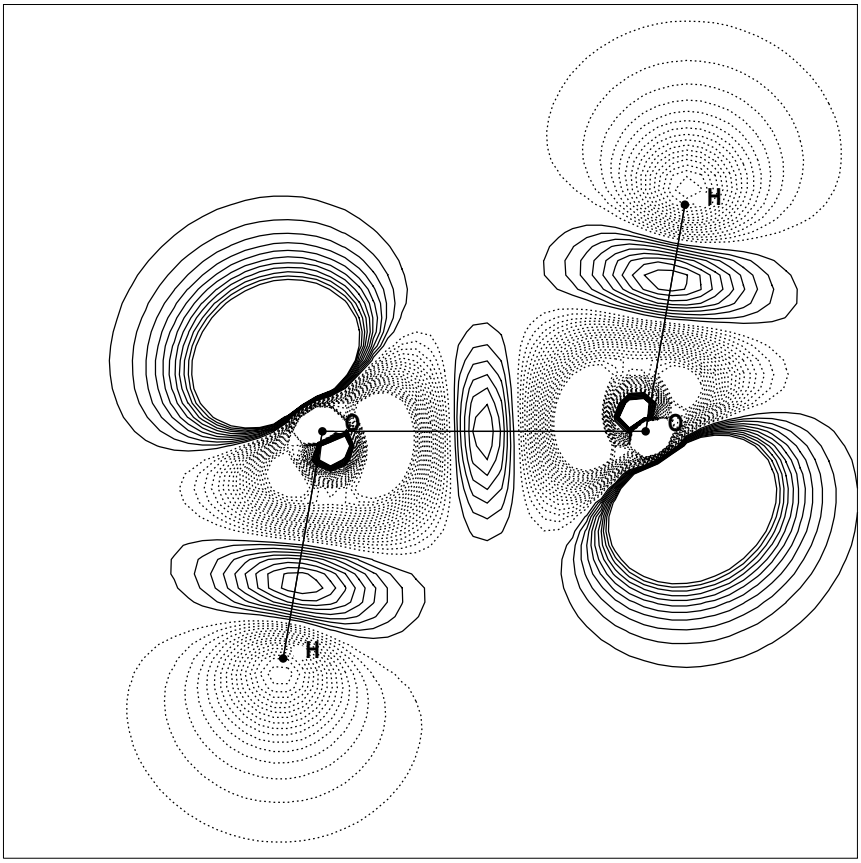


Figure 2.5: H_2O_2 sto3g Molecular density minus Oriented Oxygens, with proper treatment of directional character of atomic density, resulting in the accumulation of density (solid contours) along the O-O axis.

can provide interesting information about the chemical bond. Most computational chemistry programs have no provision to calculate it.
In Molden the *Laplacian* is calculated according to the following equations. By definition, the *Laplacian* of the electron density is the trace of the second derivative matrix of the density.

$$\nabla^2\sigma = \frac{\delta^2\sigma}{\delta x^2} + \frac{\delta^2\sigma}{\delta y^2} + \frac{\delta^2\sigma}{\delta z^2} = \sum_r \sum_s P_{rs} \left(\frac{\delta^2}{\delta x^2} + \frac{\delta^2}{\delta y^2} + \frac{\delta^2}{\delta z^2} \right) AO_r AO_s \qquad (2.10)$$

The Atomic Orbitals (*AO's*) are constructed as a linear combination of cartesian gaussian type orbitals (*gto's*)[29] $\phi_{l\,m\,n}$ with fixed coefficients and exponents α:

$$\phi_{l\,m\,n} = N_{l\,m\,n}\, x^l y^m z^n \exp -\alpha(x^2 + y^2 + z^2) \qquad (2.11)$$

$N_{l\,m\,n}$ is the normalisation constant:

$$N_{l\,m\,n} = \left[\frac{(8\alpha)^{l+m+n} l!m!n!}{(2l)!(2m)!(2n)!} \right]^{\frac{1}{2}} \left(\frac{2\alpha}{\pi} \right)^{\frac{3}{4}} \qquad (2.12)$$

The calculation of the *Laplacian* now breaks down into derivatives over products of *gto's*:

$$\frac{\delta^2\phi_i\phi_j}{\delta x^2} = \frac{\delta^2\phi_i}{\delta x^2}\phi_j + 2\frac{\delta\phi_i}{\delta x}\frac{\delta\phi_j}{\delta x} + \phi_i\frac{\delta^2\phi_j}{\delta x^2} \qquad (2.13)$$

The explicit formulas for the first and second derivatives over primitive gaussians are given by:

$$\frac{\delta\phi_{l\,m\,n}}{\delta x} = (-2\alpha x + \frac{l}{x})\,\phi_{l\,m\,n} \qquad (2.14)$$

$$\frac{\delta^2\phi_{l\,m\,n}}{\delta x^2} = \left((-2\alpha - \frac{l}{x^2}) + (-2\alpha x + \frac{l}{x})^2 \right) \phi_{l\,m\,n} \qquad (2.15)$$

Derivatives over y and z follow by analogy. The last two equations do *not* hold for $x = 0$ and in Molden this has been solved programmatically by letting x, y or z be the smallest possible number within machine precision for these cases.

The recovery of the shell structure by the *Laplacian* in carbon monoxyde is shown in Figure 2.7.

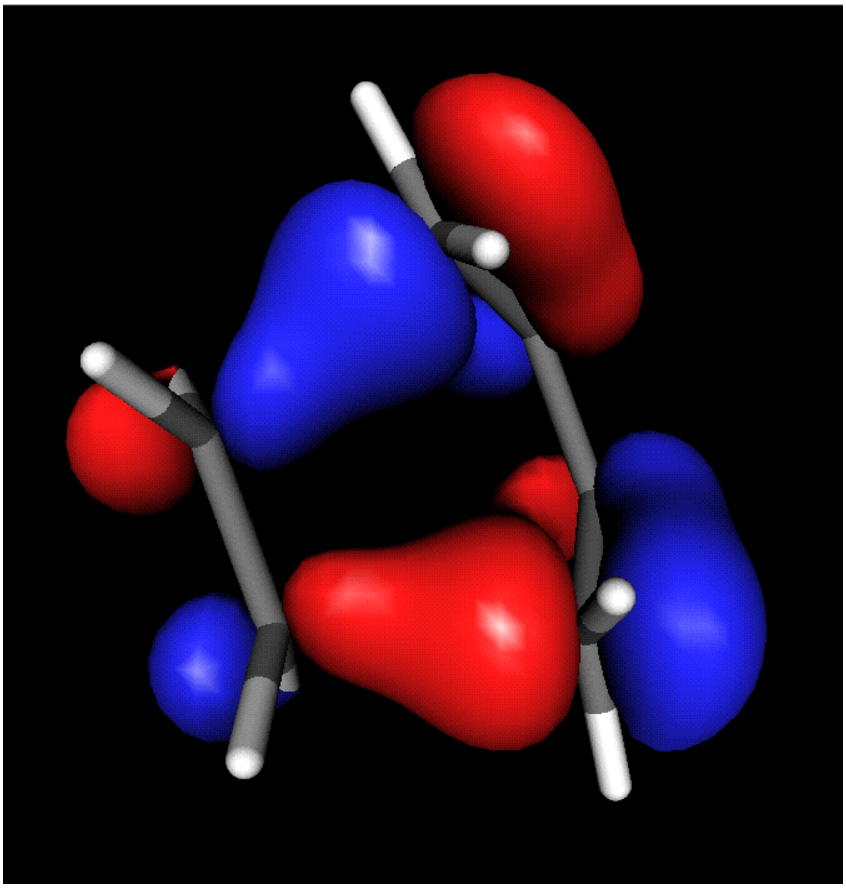


Figure 2.6: Higest Occupied Molecular Orbital of a Diels-Alder cycloaddition transition-state of butadiene with ethylene. Shown are the contour surfaces of orbital amplitude 0.25 (Blue) and -0.25 (Red).

Molecular charge distribution

Distributed Multipole Analysis

The calculation of the interaction energy of two molecules, as implemented in all force field based programs, is crucially dependent on a correct description of the

electrostatic contribution. Most force field methods describe this term as a sum of coulomb interactions between partial atomic charges. This however does not take into account the fact that an atom in the field of other atoms is polarized and exerts an electric force which is not equal in all directions. By representing the molecular charge distribution by a set of point multipoles on a number of centers (often atomic centers), the electrostatic interaction can be modeled far more accurately .[30,31]

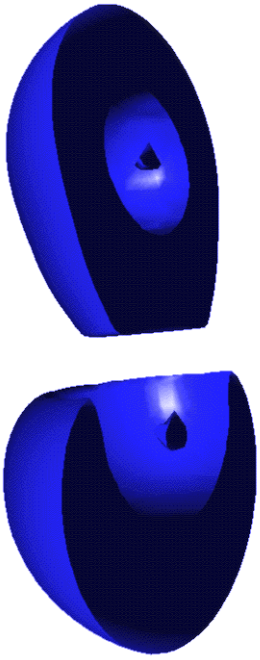


Figure 2.7: CO sto3g, surface of constant Laplacian(=-0.3), showing the 1s and valence shells of carbon and oxygen.

Such a model automatically includes the effect of lone pair and π electron density on the intermolecular forces and is widely used to model complexes of polar[32] and aromatic molecules,[33] for the prediction of protein side chain structures[34] and nucleic acid base pairs,[35] and crystal structures.[36] Calculations in the references quoted, were performed with the Orient [37] and DMAREL[38] packages, where the latter is specifically designed for crystal structure simulations.

The Distributed Multipole Analysis has been described by Stone [30,31] in a general way, but explicit formulae for the required integrals have not been reported in the literature. Molden has incorporated the calculation of multipole moments according to the following formalism.

The coulomb interaction of two charge distributions can be expressed as a

multipole expansion in the form[31]:

$$U_{elec} = (4\pi\epsilon_0)^{-1} \sum_{l_1 l_2 m_1 m_2} [l_1 l_2] R^{-(l_1+l_2+1)} Q_{l_1 m_1} Q_{l_2 m_2} S^{m_1,m_2}_{l_1,l_2,l_1+l_2} \tag{2.16}$$

Where $[l_1 l_2]$ denotes the numerical factor $[(2l_1+2l_2+1)!/(2l_1)!(2l_2)!]^{1/2}$, the Q_{lm} describe the charge distributions, and S is a function of the relative orientation of molecules. The Q_{lm} are complex for $m > 0$:

$$Q_{l,m} = (-)^m \frac{1}{\sqrt{2}}(Q_{lmc} + iQ_{lms}), \quad Q_{l,-m} = \frac{1}{\sqrt{2}}(Q_{lmc} - iQ_{lms}) \tag{2.17}$$

A molecular charge distribution can be described as set of multipole expansions at centers T, where T consists of atomic centers and centers due to interatomic overlap density. According to the gaussian product theorem,[39] the product of two gaussian type orbitals ϕ^A and ϕ^B centered at atomic positions A and B and with exponents α and β, is itself a gaussian function, centered at point T given by:

$$\overline{T} = (\alpha\overline{A} + \beta\overline{B})/(\alpha + \beta) \tag{2.18}$$

In Modern Ab Initio programs the AO's are grouped in shells, where the s,p and/or d orbitals in a shell are a fixed linear combination of the same primitive gaussians, and therefore all contribute to a given site T associated with a product of primitive gaussians. The Q_{lm} are calculated as the expectation value of the regular spherical harmonics[31]:

$$Q_{lmc}(T) = \sum_{rs} P_{rs} d_{ri} d_{sj} \langle \phi_i | R_{lmc} | \phi_j \rangle, \quad Q_{lms}(T) = \sum_{rs} P_{rs} d_{ri} d_{sj} \langle \phi_i | R_{lms} | \phi_j \rangle, \tag{2.19}$$

Here P_{rs} is an element of the density matrix, r and s run over all AO's within shell I and II respectively and each AO is expressed as a linear combination of gaussian primitive functions ϕ_i: $AO_i = \sum_r d_{ri}\phi_i$.
A multipole expansion about the point T can be represented as a multipole expansion about any other point S by means of the formula[30]:

$$Q_{lm}(S) = \sum_{k=0}^{l} \sum_{q=-k}^{k} \left[\begin{pmatrix} l+m \\ k+q \end{pmatrix} \begin{pmatrix} l-m \\ k-q \end{pmatrix} \right]^{1/2} Q_{kq}(T) R_{l-k,m-q}(S-T) \quad (2.20)$$

Thus the proliferation of expansion centers can be reduced by shifting some or all of the overlap charge distribution to the atomic centers. Molden offers three schemes for shifting the overlap charge distribution:

- shift all overlap density to the atomic sites (default)
- shift all overlap density to the nearest atomic site or to a site halfway between the bonds
- only shift overlap density between non bonded atoms

Molden calculates the multipole moments up to the hexadecapole (rank=4), since explicit terms in equation 2.16 are only available upto hexadecapole moment .[31]

We now turn to explicit formulae for the integrals of type $\langle \phi_i | R_{lmc} | \phi_j \rangle$. The regular spherical harmonics can be written as a linear combination of powers of cartesian coordinates $R = \sum_r g_r x^{l_r} y^{m_r} z^{n_r}$. and :

$$\langle \phi_i | R | \phi_j \rangle = \sum_r g_r \langle \phi_i | x^{l_r} y^{m_r} z^{n_r} | \phi_j \rangle \quad (2.21)$$

An expression for $\langle \phi_i | x^{l_r} y^{m_r} z^{n_r} | \phi_j \rangle$ can be derived in analogy for that of the overlap integral as described by Saunders[40]:

$$\langle \phi^A_{l_1 m_1 n_1} | x^{l_r} y^{m_r} z^{n_r} | \phi^B_{l_2 m_2 n_2} \rangle = \exp\left[-\alpha\beta(\overline{AB})^2/(\alpha+\beta)\right] I_x I_y I_z \quad (2.22)$$

in which:

$$\begin{aligned} I_x &= \sum_{i=0}^{l_1+l_2} f_i(l_1, l_2, \overline{TA}_x, \overline{TB}_x) \int_{-\infty}^{\infty} x^{i+l_r} e^{-(\alpha+\beta)x^2} dx \\ &= \sum_{i=0}^{l_1+l_2} f_i(l_1, l_2, \overline{TA}_x, \overline{TB}_x) \frac{(2(i+l_r)-1)!!}{(2(\alpha+\beta))^{i+l_r}} \left(\frac{\pi}{(\alpha+\beta)}\right)^{1/2} \delta(i+l_r) \end{aligned}$$

with $\delta(k) = 0$ for odd k and $\delta(k) = 1$ for even k, $\overline{TA}_x$ is the x component of the vector connecting atomic center A with the center of the gaussian product T, $\overline{TB}_x$ is the x component of the vector connecting atomic center B with the center of the gaussian product T, $(\overline{AB})^2 = (\vec{A}-\vec{B})\cdot(\vec{A}-\vec{B})$, $(2l-1)!! = 1*3*5\cdots(2l-1)$ and

$$f_k(l_1, l_2, \overline{TA}_x, \overline{TB}_x) = \sum_{i=0,l_1}^{i+j=k} \sum_{j=0,l_2} (\overline{TA})_x^{l_1-i} \begin{pmatrix} l_1 \\ i \end{pmatrix} (\overline{TB})_x^{l_2-j} \begin{pmatrix} l_2 \\ j \end{pmatrix} \qquad (2.23)$$

Expressions for I_y and I_z can be obtained by replacing l with m and n respectively and replacing x by y and z respectively. Molden can thus calculate distributed multipole moments for Ab Initio wavefunctions as read from the outputs of the beforementioned Computational Chemistry packages. These distributed multipole moments can be used to calculate an approximate electrostatic potential using the expression for the coulomb interaction of two multipole expansions described above. The distributed multipole moments are also used to generate input for the DMAREL program.

The Electrostatic Potential and ESP derived charges

Since the majority of force field packages is not capable to deal with distributed multipoles, point atomic charges are frequently used to describe the electrostatic component of the energy in force field calculations. One way to obtain these charges is by fitting an atom-centered monopole approximation to the molecular electrostatic potential, MEP. The MEP can be used to predict the sites in a molecule which are most reactive towards protonation or more generally, towards an electrophilic attack.

As first defined by Bonaccorsi et al.,[41] the MEP, V($\mathbf{r}$) represents the value, at first order of perturbation, of the interaction energy between molecule M and a proton located in $\mathbf{r}$:

$$V(\mathbf{r}) = \sum \frac{Z_a}{|\mathbf{r} - \mathbf{r_a}|} - \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' \tag{2.24}$$

The first term corresponds to nuclear repulsion from the nuclei with charge Z_a and the second originates from electronic attraction between the proton and the electron density ρ. The latter is equivalent to the Nuclear Attraction Integrals (NAI) required for the construction of the 1-electron Fock matrix in Hartree Fock calculations.[40] The algorithm currently implemented in Molden calculates the *NAI*s utilizing the Rys Polynomial Method[42,43] and a scheme developed by Besler et. al.[44] where the *MEP* is calculated on a number of Connolly surfaces, generated by scaling the Van der Waals radii of the atoms. Fitting is accomplished by a least-squares procedure. It has been noted that charges computed by such a procedure may be conformationally dependent and procedures exist which allow for the fitting of charges to multiple conformations.[45] It has also been noted[46] that sometimes not all charges can be determined with the simple least-squares procedure. By application of a singular value decomposition on the least-squares matrix an estimate of the number of charges that can be assigned with statistical validity can be obtained. Both the fitting of charges to multiple conformations and the singular value decomposition approach have not yet been used in Molden.

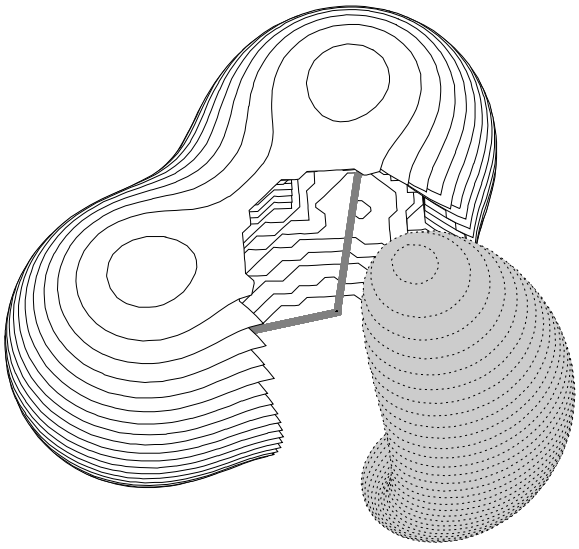


Figure 2.8: The Molecular Electrostatic Potential for the water molecule, MEP=+/- 0.1 Hartree (grey area =-0.1)

Molden calculates the *MEP* for a variety of plots; the 2D contour plot, the 3D iso-contour plot as in Figure 2.8 and color-coded on a Van der Waals surface.

2.4 Discussion

We have presented a description of the current state of development of Molden. The animation options and the treatment of difference density, the *Laplacian* of the electron density, the distributed multipoles and the electrostatic potential have been extended and where necessary theoretically underpinned. Ongoing development include the incorporation of charges fitted to the DMA derived electrostatic potential. Preliminary calculations have shown that the computational costs of the procedure are orders of magnitude less than those of the traditional ESP charge calculations. The location of extrema on the electrostatic potential map, the interfacing to the DMAREL program and visualization of the spin density are other Molden developments. Some computational chemistry programs are being interfaced to Molden, e.g. the Quantum Theory Project,[47] the MOLPRO package[48] and the ADF package[49] (Prof. A. Bencini, Inorganic Computational Chemistry, University of Florence). Molden has established a sizeable user base world wide.

Acknowledgements

The investigations reported in this paper were supported by the Netherlands Organization for Chemical Research (NWO-CW) within the framework of the PPM/CMS Crystallization project (CMS-c). The CMS-c project is a Dutch research collaboration with academic and industrial partners, focussing on precompetitive research into modeling, packing, morphology and industrial crystallization of organic compounds. Project information is a accessible at URL: http://www.caos.kun.nl/cmsc

References and Notes

[1] Frish, M.J., Trucks, G.W., Head-Gordon, M., Gill, P.M.W., Wong, M.W., Foresman, J.B., Johnson, B.G., Schlegel, H.B., Robb, M.A., Replogle, E.S., Gomperts, R., Andres, J.L., Raghavachari, K., Binkley, J.S., Gonzalez, C., Martin, R.L., Fox, D.J., Defrees, D.J., Baker, J., Stewart, J.J.P., Pople, J.A., Gaussian94. Gaussian, Inc., Carnegie Office Park, Pittsburgh, PA 15106.

[2] Schmidt, M.W., Boatz, J.A., Baldridge, K.K., Koseki, S. , Gordon, M.S., Elbert, S.T., Lam, B., GAMESS, Program No. 115. Indiana University, Bloomington, Indiana.

[3] Guest,M.F., Kendrick, J., van Lenthe, J.H., Shoeffel, K. , Sherwood, P., GAMESS-UK Users Guide and Reference Manual. Computing for Science (CFS) Ltd, Daresbury Laboratory, 1994.

[4] Stewart, J.J.P., QCPE Bull., 10 (1990) 86.

[5] Nielsen, A.B, Holder, A.J., GaussView User's Reference. Gaussian, Inc., Carnegie Office Park, Carnegie, PA 15106 USA.

[6] Tripos Associated Inc., St. Louis, USA.

[7] Molecular Simulations Inc., 9685 Scranton Road, San Diego CA, USA, Cerius User Guide, March 1997., 1997.

[8] Schaftenaar, G., QCPE Bull., 12 (1992) 3.

[9] Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E., Bryce, M.D., Rogers, J.R., Kennard, O., Shikanouchi, T., Tasumi, M., J. Mol. Bio., 112 (1977) 535.

[10] Wishart, D.S., Willard, L., Richards, F.M., Sykes, B.D., VADAR: A Comprehensive Program Suite for Protein Structural Analysis. Protein Engineering Network of Centres of Excellence, University of Alberta, 1993.

[11] Allen, F.H., Bellard, S.A., Brice, M.D., Cartwright, B.A., Doubleday, A., Higgs, H., Hummelink, T., Hummelink-Peters, B.G., Kennard, O. , Motherwell, W.D.S., Rodgers, J.R., Watson, D.G., Acta Crystallogr., Sect. B, 35 (1979) 2331.

[12] Allen, F.H., Johnson, O., Macrae, C.F., Smith, J.M. , Motherwell, W.D.S., Galloy, J.J., Watson, D.G., Rowland, R.S., Edgington, P.R., Garner, S.E., Davies, J.E., Mitchell, G.F., CSD System Documentation. Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, UK, 1992.

[13] Minnesota Supercomputer Center, Inc., Minneapolis MN, XMol Users Manual, 1994.

[14] Chemical Design Ltd, Oxon, England, Chem-X Reference Guide, Volume III, 1993.

[15] *GRAPHICS INTERCHANGE FORMAT Version 89a.* Columbus, Ohio: CompuServe Incorporated, 1990.

[16] Stewart, R.F., J. Chem. Phys., 51 (1969) 4569.

[17] Hirshfeld, F.L., Acta Crystallogr., Sect. B, 27 (1971) 769.

[18] Kunze, K.L., Hall, M.B., J. Am. Chem. Soc., 108 (1986) 5122.

[19] Schwarz, W.H.E., Ruedenberg, K., Mensching, L., J. Am. Chem. Soc., 111 (1989) 6926.

[20] F. L. Hirshfeld, *Crystallography Reviews*, vol. 2, ch. Electron density distributions in molecules. England: Gordon and Breach, 1989.

[21] Bader, R.F.W., Keaveny, I., Cade, P.E., J. Chem. Phys., 47 (1967) 3381.

[22] Schwarz, W.H.E., Valtazanos, P., Ruedenberg, K., Theor. Chim. Acta, 68 (1985) 471.

[23] Roothaan, C.C.J., Rev. Mod. Phys., 23 (1951) 69.

[24] I. Fleming, *Frontier Orbitals and Organic Chemical Reactions.* New York: Wiley, 1976.

[25] R. Bader, *Atoms in Molecules - A Quantum Theory.* Oxford: Oxford University Press, 1990.

[26] Bader, R.F.W., Popelier, P.L.A., Keith, T.A., Angewandte Chemie, Intl. Ed. Eng., 33 (1986) 620.

[27] Bader, R.F.W., Gough, K.M., Laidig, K.E., Keith, T.A., Molec. Phys., 75 (1992) 1167.

[28] Bader, R.F.W., Gillespie, R.J., MacDougall, P.J., J. Am. Chem. Soc., 110 (1988) 7239.

[29] Clementi E., Davis D.R., J. Chem. Phys., 45 (1966) 2593.

[30] Stone, A.J., Chem. Phys. Lett., 83 (1981) 233.

[31] Price, S.L., Stone, A.J., Alderton, M., Molec. Phys., 52 (1984) 987.

[32] Buckingham, A.D., Fowler, P.W., Can. J. Chem., 63 (1985) 2018.

[33] Price, S.L., Stone, A.J., J. Chem. Phys., 86 (1987) 2859.

[34] Mitchell, J.B.O., Nandi, C.L., Thornton, J.M., Price, S.L., Singh, J., Snarey, M., J. Chem. Soc., Faraday Trans., 89 (1993) 2619.

[35] Price, S.L., Lo Celso, F., Treichel, J.A., Goodfellow, J.M., J. Chem. Soc., Faraday Trans., 89 (1993) 3407.

[36] Coombes, D.S., Price, S.L., Willock, D.J., Leslie, M., J. Chem. Phys., 100 (1996) 7352.

[37] Stone, A.J., Dullweber, A., Popelier, P.L.A., Wales, D.J., Orient: a program for studying interactions between molecules, version 3.2. University of Cambridge, 1995.

[38] Willock, D.J., Price, S.L., Leslie, M., Catlow, C.R.A., J. Comp. Chem., 16 (1995) 628.

[39] Boys, S.F., Proc. R. Soc., 200 (1950) 542.

[40] V. Saunders, *Computational Techniques in Quantum Chemistry and Molecular Physics*, ch. An Introduction to Molecular Integral Evaluation. Dordrecht: Reidel, 1975.

[41] Bonaccorsi, R., Scrocco, E., Tomasi, J., J. Chem. Phys., 54 (1970) 5270.

[42] Dupuis, M., Rys, J., King, H.F., J. Chem. Phys., 65 (1976) 111.

[43] King, H.F., Dupuis, M., J. Comp. Phys., 21 (1976) 144.

[44] Besler, B.H., Merz, K.M., Kollman, P.A., J. Comp. Chem., 11 (1990) 431.

[45] Bayly, C.I., Cieplak, P., Cornell, W.D., Kollman, P.A., J. Phys. Chem., 97 (1993) 10269.

[46] Francl, M.M., Carey, C., Chirlian, L.E., J. Phys. Chem., 17 (1996) 367.

[47] URL: http://www.qtp.ufl.edu/

[48] Werner, H.J., Knowles, P.J., User's manual for MOLPRO. Insititut fur Theoretische Chemie, Universitat Stuttgart and School of Chemistry, University of Birmingham, 1996.

[49] Baerends, E.J., Ellis, D.E., Ros, P., Chem. Phys., 2 (1973) 41.

Chapter 3

The effect of isodensity surface sampling on ESP derived charges and the effect of adding bondcenters on DMA derived charges.

3.1 Summary

The effect of sampling the electrostatic potential around a molecule on the quality of electrostatic potential derived charges is investigated. In addition, the effect of the number of expansion sites in a Distributed Multipole Analysis on the quality of charges fitted to the DMA derived electrostatic potential is investigated. Sampling on constant electron density surfaces gives a better fit between the quantum mechanical potential and the potential derived from the fitted charges, compared to sampling on a Van der Waals surface composed of intersecting spheres. The fit between the electrostatic potential derived from point charges and the quantum mechanical potential becomes poorer with an increasing quality of the employed basis set. The inclusion of bondcenters into the calculations improves the fit between the Quantum Mechanical electrostatic potential and the DMA derived potential. The number of expansion sites needed for an accurate approximation of the QM electrostatic potential increases with increasing quality of the used basis set.

3.2 Introduction

Classical force field methods use Coulomb's law to describe the electrostatic interactions of molecules. This requires the use of point charges. Most force fields use

a quick calculation of point charges based on electronegativity rules .[1,2] Several methods exist to determine point charges from a Quantum Mechanical (QM) calculation. These methods can be subdivided into a class where charges are determined by some scheme that partitions the electron density over the atoms (Mulliken,[3] Bader,[4] Hirshfeld[5]) and a class where charges are optimized to reproduce the QM electrostatic potential (ESP) by employing a least-squares fit of the model (on point charges based) potential and the QM potential. From now on we will refer to these charges as QMESP charges. Methods in the latter class differ mainly by how and where the electrostatic potential in sampled in the surrounding molecular space.[6,7] These methods sometimes have problems with statistically poorly determined charges on (buried) centers. The RESP method [8] was developed to deal with this problem by allowing the simultaneous fit of charges for multiple conformations.

However, using point charges to describe electrostatic contributions neglects the fact that an atom in the field of other atoms is polarized and exerts an electric force which is not spherically symmetric. By representing the molecular charge distribution as a set of multipoles on a number of centers, the electrostatic interaction can be modeled far more accurately .[9,10] Stone [9,10] has described a generally applicable method for the determination of distributed multipole moments from electron density determined using Quantum Mechanics (QM). However, most force field methods cannot handle multipoles and rely on atomic partial charges to describe electrostatic interactions. Exceptions are the tinker package by Ponder[11] and the DMAREL program .[12] The distributed multipoles can still be useful for point charges based force fields, since the DMA derived electrostatic potential can be used to obtain DMA derived charges (DMAESP charges). The advantage of these DMAESP charges over the conventional QMESP charges is a decrease in computational cost of a two orders of magnitude. A procedure to calculate the DMAESP charges is available in our program *Molden* and has recently been developed independently by Winn et al.[13] DMAESP charges are not equivalent to the monopoles which are the lowest rank of multipoles resulting from a Distributed Multipoles Analysis. These monopoles suffer from the same deficiency as charges derived from a Mulliken population analysis [3] in which the overlap density is equipartitioned over the contributing atoms, which is a quite arbitrary choice as explained by Williams.[14] In the DMA approach, the overlap density is

shifted (see formula (20) from the previous article;[15]) to the nearest expansion site. In the standard approach this expansion site is an atomic site, which is an equally arbitrary choice as equipartitioning. The first objective of this work is to evaluate the quality of DMAESP charges in comparison with conventional QMESP charges.

The second objective of this work is to improve the quality of both ESP and DMAESP charges, by better sampling of the QM or DMA derived electrostatic potential around the molecule.

3.3 Methods

3.3.1 The quality of DMA derived charges versus QMESP charge

The agreement between of DMAESP charges and conventional QMESP charges, is limited by how well the DMA derived electrostatic potential approximates the QM electrostatic potential. From now on we will refer to the QM electrostatic potential as "true potential". Not shifting the overlap density to atomic sites should in priciple produce the best agreement between the DMA potential and the QM potential. This method is only tractable for the smallest molecules, because the number of overlap sites scales with the square of the number of primitive cartesian gaussians used to describe the electron density and thus increases with basis-set quality. It is expected that increasing the number of expansion sites will systematically improve the agreement between the true electrostatic potential and the DMA derived potential and that more expansion sites are needed as the quality of the basis-set increases. We will compare the agreement between the DMAESP charges with those of the true ESP equivalent, for two DMA variants: the standard approach with only atomic expansion sites and a variant where bondcenters have been added.

The quality of the DMA derived potential can be inspected by visualizing the difference between the QM potential and the DMA derived potential. The quality of the charges will be judged by the root mean square of the difference in atomic partial charges derived from the QM potential and the DMA derived potential (Qrms) for a test suite of molecules. The mathematical background of electrostatic potentials, the charge fitting procedure and the Distributed Multi-

pole Analysis are covered earlier in some detail .[15] The 2-(methylammonium)-propanol salt of a cyclic phosphoric acid (the so-called "pa-complex") has highly localized charges; the phosphorus is strongly positive and two of the oxygens bonded to it carry a nearly full negative charge. An additional positive charge is located on the nitrogen atom and the complex has two hydrogen bonds, see Figure 3.1. For an accurate QM calculation diffuse functions on the negatively charged oxygens are required. This complex will serve as a test case for a highly inhomogeneous electrostatic potential.

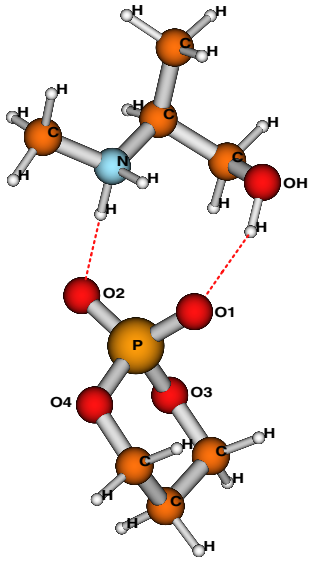


Figure 3.1: Test suite "pa-complex" with highly localized charges.

3.3.2 The effect of sampling on the quality of DMA derived charges and QMESP charges

The second objective of this work is an attempt to improve the quality of both ESP and DMAESP charges, by better sampling of the QM or DMA derived electrostatic potential around the molecule. The quality criterion in all ESP

methods is a minimal quadratic sum of the deviation of the model potential from the QM electrostatic potential at the points used in the fit. Henceforth we will refer to it as the goodness-of-fit or GOF:

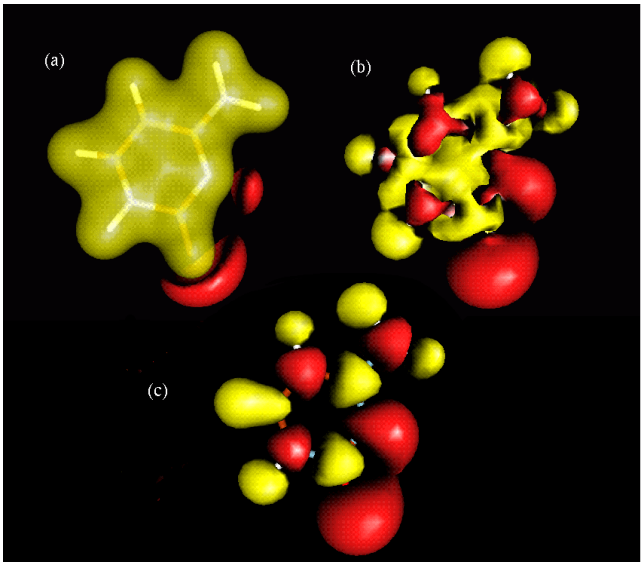


Figure 3.2: Isopotential surfaces (potential=+0.01 Hartree (yellow), -0.01 Hartree (red)) of the cytosine molecule for three types of potentials; (a) the QM electrostatic potential, (b) the DMA derived potential and (c) the ESP charges derived potential

$$GOF = \sqrt{\frac{\sum_i^m (V_i^o - V_i^c)^2}{n}}$$

where V_i^o is the QM electric potential and V_i^c the model potential which is given by Coulomb's law:

$$V_i^c = \sum_j^n \frac{q_j}{r_{ij}}$$

where i runs over the points sampling the electric potential (m in total) and j runs over the number of charges to be fitted (n in total). r_{ij} is the distance

between point charge j and sampling point i. Large potentials and large deviations tend to be strongly expressed in the GOF giving unfairly great weights to the points nearby atoms. This may ultimately result in a poor fit of the potential at the sampling points with small potentials, and a decrease in quality of the charges associated with it. It is therefore crucial for the quality of the fit not to oversample large potential values. Williams experimented with the limit of the inner boundary of the Van der Waals surface. The fit to the QM potential (rms and rrms) became worse at the smaller boundary and better at the larger boundary.[16] Singh and Kollman[17] located potential grid points in equally spaced shells located at 1.4, 1.6, 1.8 and 2.0 times the Van der Waals radii. However, using such a Van der Waals surface one uses the same Van der Waals radii regardless of the different atomic environments. In contrast to a neutral oxygen, a negatively charged oxygen atom is likely to have an electrondensity distribution that extends further into space, which corresponds with a slightly larger Van der Waals radius. Using a smaller radius will result in sampling points with a higher electron density with respect to the neutral oxygen. This in turn could lead to oversampling of high ESP values for some of the oxgygen atoms with respect to others. The magnitude of this effect will increase with increasing level of detail of the basis-set, since a more flexible basis-set allows the atoms to extend further into space.

To avoid these problems we propose to sample the electrostatic potential on a number of surfaces with constant electron density or isodensity surfaces. This has the advantage of minimizing electron-cloud penetration effects. As two molecules approach each other their potentials are distorted by electron-cloud penetration effects.[18] Using isodensity surfaces ensures that local differences in electron-cloud penetration are small in the area where the electrostatic potential is sampled. The values of the electron density are chosen such that the isodensity surfaces approach best the Van der Waals shells employed by Singh and Kollman. Sampling also effects the quality of DMAESP charges as compared to QMESP charges. The QM potential and the DMA derived potential (and the ESP charges derived potential) behave differently near the atoms (See Figure 3.2). When r (the distance to a particular atom) is small the QM potential is a function of the nuclear charge divided by r whereas a charges derived potential and the DMA derived potential, in first approximation, are a function of partial charge divided by

r. Sampling too close to the atoms will therefore reduce the quality of DMAESP charges. As discussed above, both the quality of the QMESP/DMAESP charges and the correspondence between QMESP and the DMAESP are dependent on the quality of the basis-set used; therefore we will present the basis-set dependency as a separate topic. QMESP charges mentioned in this work were all calculated using the scheme by Singh and Kollman.[17]

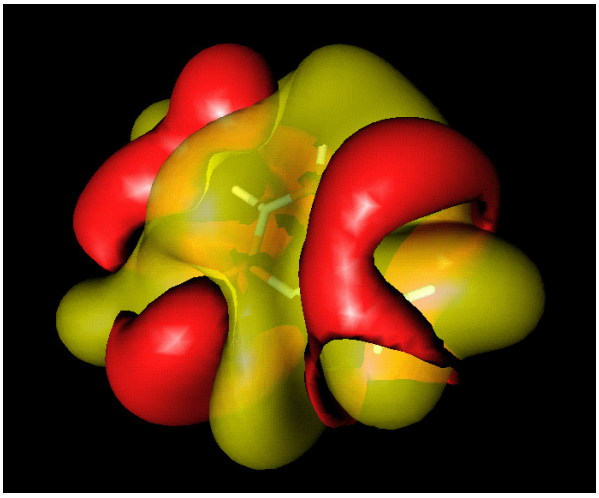


Figure 3.3: Graphical illustration of the difference between the QM electrostatic potential and the DMA potential with only atomic expansion sites for the cytosine molecule. Shown as surfaces with constant potential v. +v = yellow and -v = red. v=0.0001 Hartree (0.06 kcal/mol)

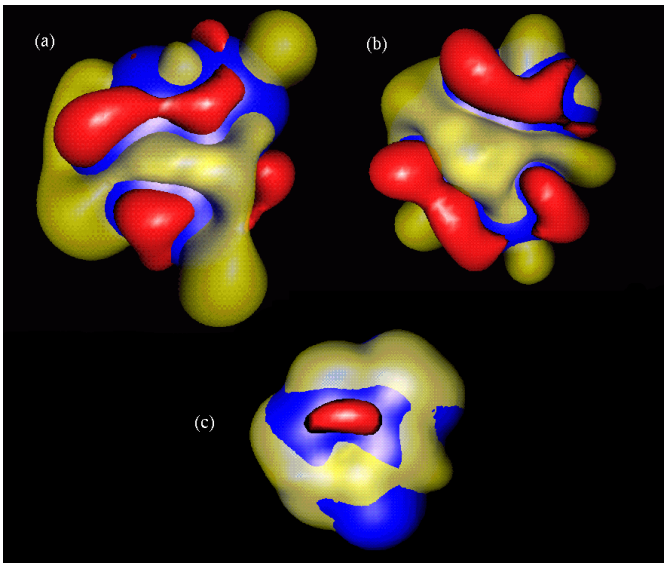


Figure 3.4: The difference of the QM electrostatic potential and the QMESP charges derived potential (a), the DMA potential derived with atomic expansion sites only (b) and the DMA potential derived with atomic and bond centers expansion sites (c). Shown as surfaces with constant function values +/- v; +v = yellow and -v = red. v=0.0001 Hartree (0.06 kcal/mol) for figures (b) and (c), v=0.0005 Hartree (0.31 kcal/mol) for figure (a). Also shown is the surface with constant density value 0.0001 $e/Bohr^3$ (blue color).

3.4 Results

3.4.1 DMA derived charges

In Table 3.1 two variants of DMAESP charges are compared with the true QMESP charges (charge set 1), using the root mean square differences of the atomic partial charges (Qrms) as criterion. In the first variant, only the atoms were used as expansion sites (charge set 2). In the second variant, in addition bondcenters were included as expansion sites (charge set 3). Both Van der Waals surface sampling and the isodensity surface sampling were performed. They are

designated by VDW and ISO respectively.

Table 3.1: Comparison of the root mean square (Qrms) of the difference between DMAESP charges and QMESP charges for two types of electrostatic potential sampling; van de Waals surface sampling (VDW) and isodensity surface sampling. (1) QMESP charges (2) DMA with atomic expansion sites (3) DMA with bondcenter sites added

compound	$Qrms_{1-2}^{VDW}$	$Qrms_{1-3}^{VDW}$	$Qrms_{1-2}^{ISO}$	$Qrms_{1-3}^{ISO}$
H_2O^{++}	0.0002	0.00004	0.0007	0.0005
$tetrahydrofuran^{++}$	0.0202	0.0128	0.0111	0.0023
$2-metoxymethyl,$ $5-metoxytetrahydrofuran^{+}$	0.0264	0.0120	0.0155	0.0032
$cytosine^{+}$	0.0164	0.0028	0.0105	0.0011
$pa-complex^{+}$	0.0269	0.0126	0.0187	0.0042
$benzene^{+}$	0.0061	0.0014	0.0031	0.0015

$^{+}$ 3-21g
$^{++}$ 6-31g**

A comparison of the charges calculated with the two DMA variants shows that the addition of bondcenters (charge set 3) to only atomic expansion sites (charge set 2) reduces the Qrms values for both VDW surface sampling and isodensity surface sampling. The difference in Qrms values is about a factor 2 for VDW surface sampling and about a factor 5 for isodensity surface sampling, clearly showing the improved charges resulting from the isodensity surface sampling.

In Figure 3.3 the overall difference between the QM potential and the DMA potential with atomic expansion sites only is illustrated graphically for the cytosine molecule. This difference is positive (yellow) near the atoms and negative (red) between the atoms, i.e. the region of space closest to the bondcenters.

This observation suggested that the agreement between both potentials could be improved by adding expansion sites at the bond centers.

The improvement of the fit is presented graphically in Figure 3.4, which shows the difference function between the QM electrostatic potential and three derived potentials:

- (a) The electrostatic potential derived from QMESP charges
- (b) The DMA derived potential with atomic expansion sites only
- (c) The DMA derived potential with atomic and bond centers expansion sites

Figure 3.4 clearly shows that the error resulting from sampling the space outside the isodensity surface is larger for the DMA derived potential with only atomic expansion sites (b) than for the potential derived with bondcenter expansion sites added (c). The error surfaces extend further into space for smaller absolute values of the error v. This can be understood if one considers that close to the atomic sites the difference function always becomes strongly positive because of the different behaviour of the QM potential and the derived potentials in this region, as discussed above. It can also be seen that the DMA derived potentials are in far better agreement with the QM potential than the QMESP charges derived potential.

3.4.2 Isodensity surface sampling vs. Van der Waals surface sampling

For the "pa-complex" in our test suite, distances of different oxygen atoms to the surface with density value 0.0001 $e/Bohr^3$, are given in Table 3.2. It can be seen that the extension in space varies up to 0.15 Ångstrom from a hydroxyl oxygen (OH), to the negatively charged oxygen bonded to phosphorus (O1).

For the same compound this isodensity surface for two different basis-sets (3-21g and 6-31g*+) and a Van der Waals surface of intersecting spheres, scaled by a factor of 1.4, are shown in Figure 3.5. This Van der Waals surface approximates the isodensity surface quite well for the medium sized basis-set but it lays well inside the isodensity surface of the bigger basis-set.

Van der Waals surface sampling will include more higher density points for the medium size basis-set than for the larger basis-set. This is equivalent to sampling

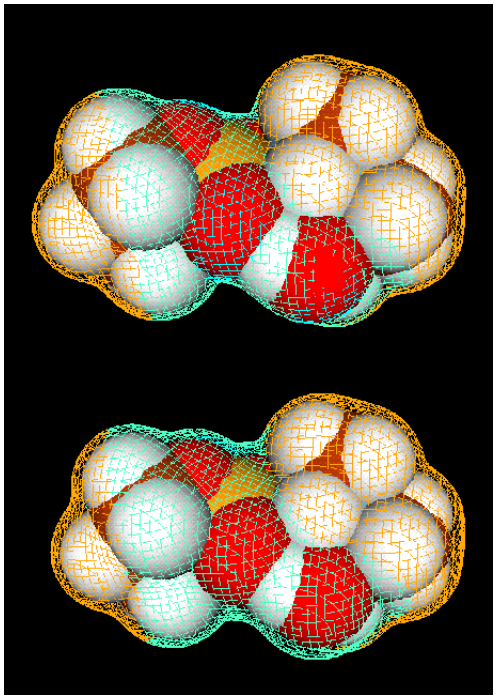


Figure 3.5: Isodensity surface with density value 0.0001 $e/Bohr^3$ for two different basis-sets (3-21g (top) and 6-31g+ (bottom)). The electrostatic potential is color coded on the surface (positive values orange, negative values blue). The Van der Waals surface of intersecting spheres with a Van der Waals radius scaled by a factor 1.4 is shown as a solid surface.*

Table 3.2: Distance to the surface with constant density for different oxygen atoms in the "pa-complex"

Atom	Distance (in Ångstroms) to isodensity surface (density=0.0001 $e/Bohr^3$)
OH	2.17
$O1$	2.32
$O2$	2.28
$O3$	2.22
$O4$	2.22

more points with a lower electrostatic potential, since higher electron density means more shielding of the nuclear charge. Only for the medium quality basis-set, isodensity surfaces at density values 0.0001, 0.00001 and 0.000001 $e/Bohr^3$ are fairly well approximated by Van der Waals surfaces located at 1.4, 1.6 and 1.8 times the Van der Waals radii respectively.

Comparison of the goodness-of-fit between VDW surface sampling and isodensity surface sampling (See table 3.3) shows that the isodensity surface sampling yields significantly better goodness-of-fit values. There is also a general trend towards better correspondence with the QM dipole moment when isodensity surface sampling is used, although this trend is not as strong as with the GOF values.

Table 3.3: Comparison of the Van der Waals surface sampling (VDW) and the isodensity surface sampling (ISO), based on the GOF's (in kcal/mol) and the differences (ΔD) between the dipolemoment based on charges and the quantum-mechanical dipole moment (D), both in Debye

Compound	GOF^{VDW}	GOF^{ISO}	ΔD^{VDW}	ΔD^{ISO}	D
H_2O^{++}	1.5979	1.4567	0.0458	0.0434	2.1475
$tetrahydrofuran^{++}$	1.2738	0.8919	−0.0261	0.0036	1.9372
$2-metoxymethyl,$	1.0034	0.7873	−0.0021	0.0174	4.6624
$5-metoxytetrahydrofuran^{+}$					
$cytosine^{+}$	0.7394	0.6064	0.0491	0.0200	7.1521
$pa-complex^{+}$	0.9891	0.7784	0.0145	−0.0009	4.3139
$benzene^{+}$	0.9940	0.5128	0.0025	0.0020	0.0000

$^{+}$ 3-21g
$^{++}$ 6-31g**

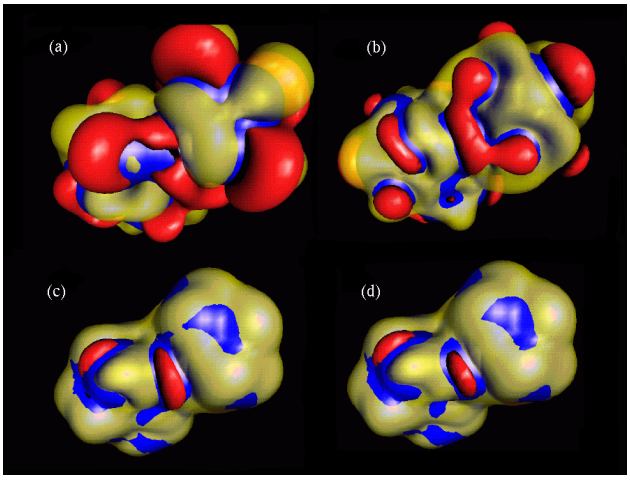


Figure 3.6: The difference function of the QM electrostatic potential (3-21g basis-set) and (a) QMESP charges derived potential (b), (c), (d) three variants of the DMA derived potential. Shown as surfaces with constant function value v for the "pa-complex". +v = yellow and -v = red. v=0.0001 Hartree (0.06 kcal/mol) for (b),(c),(d) and v=0.0005 Hartree (0.31 kcal/mol) for (a). the surface with constant density value 0.0001 $e/Bohr^3$ is shown in blue

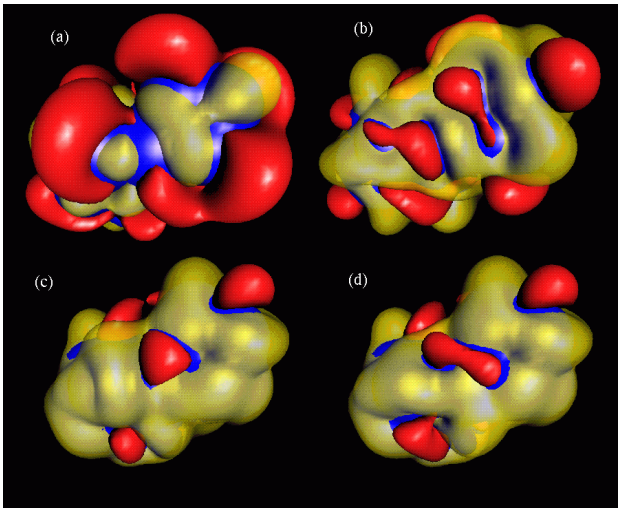


Figure 3.7: The difference function of the QM electrostatic potential (6-31g*+ basis-set) and (a) QMESP charges derived potential (b), (c), (d) three variants of the DMA derived potential. Shown as surfaces with constant function value v for the "pa-complex". +v = yellow and -v = red. v=0.0001 Hartree (0.06 kcal/mol) for (b),(c),(d) and v=0.0005 Hartree (0.31 kcal/mol) for (a). the surface with constant density value 0.0001 $e/Bohr^3$ is shown in blue

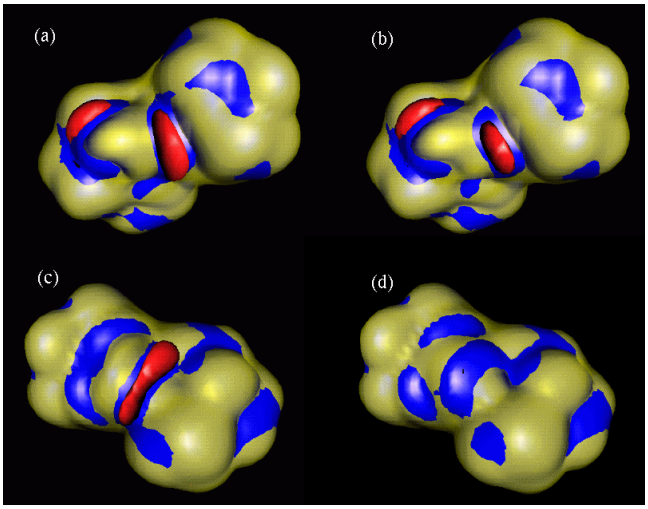


Figure 3.8: The difference function of the QM electrostatic potential (3-21g basis-set) and DMA derived potential; (a) front-side view and (c) back-side view of "pa-complex" for bond center sites only. (b) and (d) idem with hydrogenbond center sites added. Shown are surfaces with constant function value v, +v yellow and -v red. v=0.0001 Hartree (0.06 kcal/mol). The surface with constant density value 0.0001 $e/Bohr^3$ is shown in blue.

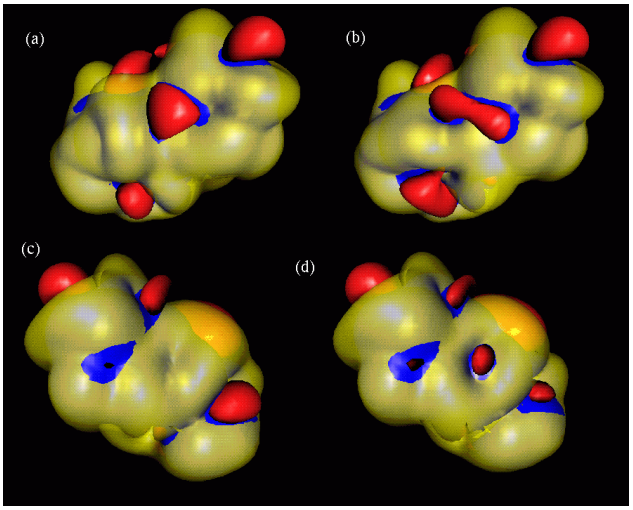


Figure 3.9: The difference function of the QM electrostatic potential (6-31g*+ basis-set) and DMA derived potential; (a) front-side view and (c) back-side view of "pa-complex" for bond center sites only. (b) and (d) idem with hydrogenbond center sites added. Shown are surfaces with constant function value v, +v yellow and -v red. v=0.0001 Hartree (0.06 kcal/mol). The surface with constant density value 0.0001 $e/Bohr^3$ is shown in blue.

Basis-set quality influences the fit between the QM electrostatic potential and derived potentials. This effect is clearly shown by the GOF values presented in Table 3.4. The effect is most pronounced for Van der Waals surface sampling. There the GOF becomes significantly worse with increasing basis-set quality. For isodensity surface sampling the GOF is not only better but also does not significantly increase with increasing basis-set quality.

Table 3.4: The GOF (in kcal/mol) as a function of the basis-set quality for Van der Waals surface sampling (VDW) and isodensity surface sampling (ISO). Data presented are for the "pa-complex".

$basis-set$	GOF^{VDW}	GOF^{ISO}
$sto3g$	0.8237	0.6904
$3-21g$	0.9891	0.7784
$6-31g**$	1.0905	0.6275

Besides effecting the fit, basis-set quality also effects the reliability of the DMAESP charges. Independent of sampling method and the DMA variant used, the deviation of DMAESP charges from QMESP charges deteriorates with increasing basis-set quality as shown by the Qrms values in Table 3.5.

The larger the basis set one selects to calculate the QM electrostatic potential, the less is the fit with derived potentials and the less charges can be approximated by any less compute intensive method. However, by addition of more expansion sites to the DMA analysis, DMA derived potentials can be systematically improved.

The deteriorating agreement with increasing basis-set flexibility between QM electrostatic potential and derived potentials and the improvement in fit by adding additional expansion sites is illustrated graphically in Figures 3.6, 3.7, 3.8 and 3.9. The decrease in fit is understood by considering that the ratio atomic plus overlap sites versus atomic plus bond center sites becomes increas-

Table 3.5: Root mean square values (Qrms) of the deviation of DMAESP charges from QMESP charges as a function of basis-set quality and sampling method, Van der Waals surface sampling (VDW) and isodensity surface sampling (ISO). Data for the "pa-complex". (1) QMESP charges (2) DMA with atomic expansion sites (3) DMA with bondcenter sites added

$basis-set$	$Qrms_{1-2}^{VDW}$	$Qrms_{1-3}^{VDW}$	$Qrms_{1-2}^{ISO}$	$Qrms_{1-3}^{ISO}$
$sto3g$	0.0160	0.0084	0.0096	0.0033
$3-21g$	0.0269	0.0126	0.0187	0.0042
$6-31g*+$	0.0347	0.0282	0.0255	0.0181

ingly unfavorable with increasing flexibility of the basis-set used. Figures 3.6 (3-21g basis-set) and 3.7 (6-31g*+ basis-set) show the difference function between the QM electrostatic potential and four derived potentials for the molecular "pa-complex":

- (a) The electrostatic potential derived from ESP charges
- (b) The DMA derived potential with atomic expansion sites only
- (c) The DMA derived potential with atomic and bond centers expansion sites
- (d) The DMA derived potential with atomic, bond center and hydrogen-bond center expansion

Comparison of Figures 3.6 and 3.7 shows the basis-set effect; the systematic improvement as a result of the addition of expansion sites is clearly illustrated in going from (a) to (d) in both Figure 3.6 and 3.7.

The improvement of the fit between the QM electrostatic potential and DMA derived potentials as an effect of the addition of hydrogenbond center expansion

sites on top of atomic and bond center sites is shown in more detail in Figures 3.8 and 3.9 where we zoom in on potentials (c) and (d) above.

The effect of the addition of hydrogen-bond center expansion sites on the quality of the fit is most significant at the back side of the molecule where the spacial extension of electron density is less pronounced. The locations where the DMA derived potential still deviates from the QM potential could be identified as oxygen lone pair sites.

3.4.3 Timings

Table 3.6 shows a comparison of times needed to calculate QMESP charges (t_{QMESP}) and DMAESP charges (t_{DMAESP}) as a function of the number of basis functions used to describe the QM electron density. The ratio t_{QMESP} / t_{DMAESP} increases in favor of the DMAESP charges with increasing number of basis functions, ranging from 16 for small systems to 56 for larger systems.

Table 3.6: Comparison of the calculation time of QMESP charges (t_{QMESP}) and DMAESP charges (t_{DMAESP}) as a function of number of basis functions used. Timings are in seconds

Compound	t_{QMESP}	t_{DMAESP}	t_{QMESP}/t_{DMAESP}	*Basis Functions*
H_2O^{++}	4.28	0.26	16.5	25
$benzene^{+}$	53.2	1.4	38.0	66
$cytosine^{+}$	84.5	2.0	42.3	82
$tetrahydrofuran^{++}$	165.5	3.7	44.7	115
$2-metoxymethyl,$ $5-metoxytetrahydrofuran^{+}$	253.0	5.4	46.9	118
$pa-complex^{+}$	619.3	11.0	56.3	166

$^{+}$ 3-21g
$^{++}$ 6-31g**

3.5 Conclusions

Calculations on a set of test molecules with wide variety of charge distributions were performed to establish the quality of the DMA derived electrostatic potentials and the effect of the sampling method on the quality of charges fitted to a DMA derived potential. The advantage of DMAESP charges over the conventional (QMESP) charges is a decrease in computational cost of a few orders of magnitude. It was found that the quality of the DMA derived electrostatic potential can be systematically improved by adding additional expansion sites to the DMA analysis and that more sites are needed with increasing basis-set quality. The same procedure improves the DMAESP charges as well. Sampling the electrostatic potential on isodensity surfaces instead of sampling on a Van der Waals surfaces improves both ESP and DMAESP charges. Using the

isodensity sampling method and a sufficiently large number of expansion sites, DMAESP charges can be calculated in good agreement with the conventional QMESP charges, with typical errors being smaller than one percent.

3.6 Acknowledgements

The investigations reported in this paper were supported by the Netherlands Organization for Chemical Research (NWO-CW) within the framework of the PPM/CMS Crystallization project (CMS-c). The CMS-c project is a Dutch research collaboration with academic and industrial partners, focussing on pre competitive research into modeling, packing, morphology and industrial crystallization of organic compounds. Project information is a accessible at URL: http://www.caos.kun.nl/cmsc.

References and Notes

[1] Del Re, G., J. Chem. Soc. London, 1958 (1958) 4031.

[2] Gasteiger, J., Marsili, M., Tetrahedron, 36 (1980) 3219.

[3] Mulliken, R.S., J. Chem. Phys., 23 (1955) 1833.

[4] R. Bader, *Atoms in Molecules - A Quantum Theory.* Oxford: Oxford University Press, 1990.

[5] Hirshfeld, F.L., Theor. Chim. Acta, 49 (1977) 129.

[6] Breneman, C.M., Wiberg, K.B., J. Comp. Chem., 11 (1990) 361.

[7] Besler, B.H., Merz, K.M., Kollman, P.A., J. Comp. Chem., 11 (1990) 431.

[8] Bayly, C.I., Cieplak, P., Cornell, W.D., Kollman, P.A., J. Phys. Chem., 97 (1993) 10269.

[9] Stone, A.J., Chem. Phys. Lett., 83 (1981) 233.

[10] Price, S.L., Stone, A.J., Alderton, M., Molec. Phys., 52 (1984) 987.

[11] Ponder, J. W., Richards, F. M., J. Comp. Chem., 8 (1987) 1016-1024.

[12] Willock, D.J., Price, S.L., Leslie, M., Catlow, C.R.A., J. Comp. Chem., 16 (1995) 628.

[13] Winn, P.J., Ferenczy, G.G., Reynolds, C.A., J. Phys. Chem., 101 (1997) 5437.

[14] D. E. Williams, *Reviews in Computational Chemistry*, ch. Net Atomic Charge and Multipole Models for the ab initio Molecular Electric Potential. New York: VCH Publishers, Inc., 1991.

[15] Schaftenaar, G., Noordik, J.H., J.Comp.-Aided Molecular Design, accepted for publication

[16] Cox, S.R., Williams, D.E., J. Comp. Chem., 2 (1981) 304.

[17] Singh, U.C., Kollman, P.A., J. Comp. Chem., 5 (1984) 129.

[18] Colonna, F., Evleth, E., J. Comp. Chem., 13 (1992) 1324.

Chapter 4

Quantum mechanical and force field calculations on the diastereomeric salts of cyclic phosphoric acids with ephedrine

4.1 Summary

We have calculated the relative lattice energies of the diastereomers of cyclic phosphoric acid and its chlorine derivative with ephedrine, with various computational models and compared them with experimental data. All computational models gave good structural agreement with the experiment, but only some models reproduced the experimental stability order. Calculations with the DREIDING force field in combination with several charge sets failed to reproduce the experimental stability order of the diastereomers. By using distributed multipoles to model the electrostatic interactions in force field calculations, the correct stability order was reproduced in one case, but the results are very sensitive to the factor used for scaling the electrostatic interactions and to conformational energy corrections of the rigid bodies. Quantum mechanical calculations with the DMol^3 density functional package predict a correct stability order for one pair of diastereomers, but fail to position a third polymorph correctly. Similar calculations with the VASP and SIESTA density functional packages predict the stability order of all diastereomers correctly at the Γ point in k-space. With better k-space sampling, however, the agreement between theory and experiment becomes less. An experimentally unknown chlorine-free analogue of the experimentally most stable chlorine-containing diastereomer, was calculated to be the least stable. Model systems of the two types of hydrogen-bonded chains observed in a series of diastereomeric salts with screw and translational symmetry were optimized at the Hartree-Fock and density functional level. The experimentally most frequently

occurring hydrogen-bonded chain with screw axis symmetry was calculated to be the most stable.

4.2 Introduction

During the synthesis of optically active compounds one often obtains a racemic mixture of stereoisomers. These stereoisomers or enantiomers may have very different biological activities. In the worst case the undesired enantiomer may be highly toxic or even teratogenic. The Softenon affair[1] was a dramatic example of this. One way of purifying a racemic mixture is to add a so-called resolving agent to the mixture. The resolving agent is necessarily also an optically active compound, that forms diastereomeric salts with the enantiomers of the racemic mixture. A good resolving agent yields diastereomeric salts with very different solubilities. This method is known as selective crystallization of diastereomers. In his Ph.D. thesis Leusen[2] proved that chiral discrimination is not a solution determined property, but manifests itself mainly in the solid state. He established a model that links the resolution efficiency to the crystal lattice energy difference between diastereomeric salt pairs.[3] Leusen tried to model the lattice energy differences between five diastereomeric salt pairs of ephedrine[a] and a cyclic phosphoric acid[b] with molecular mechanics force field calculations.[4] A good agreement between calculated and experimental structures was found. The calculated energy differences however, were not in agreement with experiment and in some cases the stability order was predicted incorrectly. He concluded that the description of the electrostatic interactions within the force field was insufficiently accurate and that the inclusion of polarization terms in the force field is essential. Whilst Leusen worked on a system with highly localized charges, Hansen performed similar calculations on a much wider range of compounds including diastereomeric salts.[5] She concluded that force fields are generally not very good at reproducing the relative stability order of polymorphs. Limitations of the force fields are their inabilities to handle crystals of compounds containing phosphorous or sulpher as used by Leusen. Hansen also concluded that it is not generally possible to show that a particular charge set is superior, i.e., she showed that a given charge

[a] α-[(1-methylamino)ethyl] benzyl alcohol

[b] 5,5-dimethyl-2-hydroxy-4-phenyl-1,3,2-dioxaphosphorinane 2-oxide

scheme performed well for some structures and badly for others. From the above it is clear that an improvement in the description of the electrostatic interactions would be crucial for improving the reproduction of the relative stability order of diastereomeric salts and polymorphs. The aim of our work is to find such better computational methods.
This may be accomplished in a number of ways:

- using force field calculations with improved models of point charges by:
  - modelling charges according to quantum mechanical derived electrostatic potentials of the anion-cation complexes of the diastereomeric salts instead of the ions alone. In this way we intend to incorporate some polarization of the anion by the cation and vice versa.
  - doing the same for multiple conformations of the ions and/or the anion-cation complex.
  - improving the quality of the models of the charges by better sampling of the quantum mechanical electrostatic potential. This is accomplished by sampling the potential on isodensity surfaces as opposed to the often-used used sampling on a van der Waals surface.
- employing force field calculations with distributed multipoles instead of point charges to describe the electrostatic interactions.
- using quantum mechanical codes that can deal with periodic boundary conditions appropriate for a crystalline lattice.

We evaluated these methods on a subset of two of the diastereomeric salt pairs used by Leusen. These are the salt pairs of ephedrine with cyclic phosphoric acid (CPA) and ephedrine with chlorine-substituted cyclic phosphoric acid (ClCPA). Accurate experimental geometries and relative lattice energies are available for these systems.[2]

The second objective of our work is to understand the relative stabilities of subsystems of the different crystal packings of these diastereomeric salts, in particular the two different types of hydrogen bonded chains found in these crystals.

4.3 Methods

4.3.1 The crystal structures of cyclic phosphoric acid derivatives and ephedrine

We have studied a subset of two of the diastereomeric salt pairs used by Leusen. These are the diastereomeric salts pairs of cyclic phosphoric acid with ephedrine (CPA-E) and chlorine-substituted cyclic phosphoric acid with ephedrine (ClCPA-E). The reason for limiting ourselves to these structures, is the computational cost of the quantum mechanical calculations involved. These two particular pairs of diastereomeric salts were chosen on the basis of extreme differences in lattice energies. The highest difference in experimental lattice energy was found for the salt pair of chlorosubstituted cyclic phosphoric acid (about 2 kcal/mol) and lowest (zero within experimental error) for unsubstituted cyclic phosphoric acid. Whilst the first stands the best chance of being outside the error range inherent in the different computational methodologies, the second can be used as a null-calibration. Table 4.1 shows the absolute and relative experimental heats of formation and the entropy-corrected lattice enthalpy difference ΔH_{solid}^{corr} as defined by Leusen for the diastereomers of CPA-E and ClCPA-E. Throughout this article we use the ΔH_{solid}^{corr} data for comparison with calculated lattice energy differences.

The structures of diastereomeric salts pair CPA-E have the reference codes FIMVEC and FILGAI in the Cambridge Structural Database (CSD).[6] The structures of diastereomeric salts pair ClCPA-E have the CSD reference codes FIMTUQ and FIMVAY. Table 4.1 includes the names of the salts used by Leusen et al.

The structure SUMWEC is a polymorph of the FIMVAY structure .[7] It was discovered after the publication of Frank Leusen's Ph.D. thesis. No solubility data have been published for this polymorph, so its ΔH_{solid}^{corr} value can not be calculated. The difference between the relative heats of formation and ΔH_{solid}^{corr} is very small for the diastereomers FILGAI and FIMVEC. So we have used the relative heat of formation for the structurally very similar SUMWEC. Figure 4.1 shows deprotonated CPA and protonated ephedrine, the forms in which they are present in the diastereomeric salts, with a rough indication of the charge centres.

Figure 4.2 shows the experimental structures of the diastereomers of CPA (FIMVEC FILGAI) and ClCPA (FIMTUQ SUMWEC FIMVAY) with ephedrine.

Table 4.1: Absolute and relative experimental heats of formation and the entropy-corrected lattice enthalpy difference ΔH_{solid}^{corr} (in kcal/mol) of the diastereomers of CPA-E and ClCPA-E

	Name	*refcode*	ΔH_f	$-\Delta\Delta H_f$	ΔH_{solid}^{corr}
CPA-E					
	INAP	*FIMVEC*	11.4 ± 0.3	0	0
	INAM	*FILGAI*	11.6 ± 0.1	-0.2	-0.1
ClCPA-E					
	CLINAP	*FIMTUQ*	9.4 ± 0.2	0	0
	CLAM2	*SUMWEC*	9.5	-0.1	
	CLINAM	*FIMVAY*	11.4 ± 0.2	-2.0	-2.8

The diastereomeric salt pairs are structurally similar. This can be seen in Figure 4.3, which shows the match of the salt pair of the unsubstituted CPA (FIMVEC-FILGAI) and the match of the salt pair of ClCPA (FIMTUQ-SUMWEC). All five crystal structures adopt the $P2_1$ space group, characterized by a two-fold screw-axis along the crystallographic b-axis. Apparently this crystal packing can accommodate both enantiomers of ephedrine without too much strain.

There is also a strong structural similarity between the unsubstituted and chlorosubstituted CPA salts. This can be seen in Figure 4.4, which shows the match between the unsubstituted cyclic phosphoric acid salt FIMVEC with its chlorine derivative counterpart FIMTUQ and the match between the unsubstituted cyclic phosphoric acid salt FILGAI and its chlorine derivative counterpart

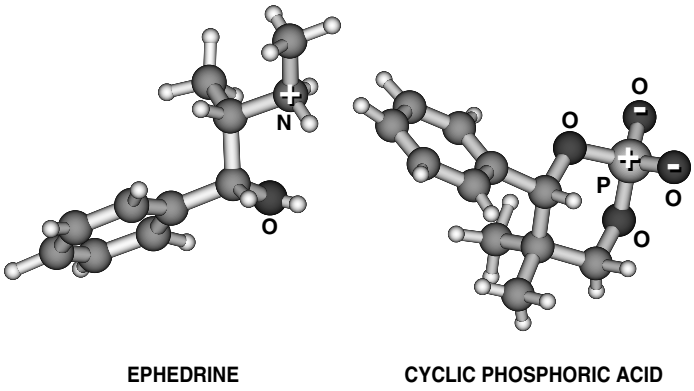


Figure 4.1: Cyclic phosphoric acid and ephedrine in protonated and deprotonated forms, with a rough indication of the charge centres.

SUMWEC. This shows that the larger chlorine can be accommodated with little strain.

These four crystal structures all adopt the same crystal packing. It is characterized by a screw-symmetric hydrogen bonded chain, as shown in Figure 4.5. This crystal packing is adopted in eight structures studied by Leusen.

The most stable chlorosubstituted CPA salt (FIMVAY) adopts a very different crystal packing, characterized by hydrogen bonded chains with translational symmetry, as shown in Figure 4.6.

Figures 4.7 and 4.8 show simplified forms of the hydrogen bonded chains with translational and screw symmetry, respectively. NO1 and NO2 are defined as the vectors from the ephedrine nitrogen hydrogen-bond donor to the phosphoric acid oxygen hydrogen-bond acceptor. With screw symmetry, the two vectors are distinct. Whilst NO2 runs roughly parallel with the chain, NO1 is more or less perpendicular to the chain direction. With translational symmetry the difference between the two is not very distinct.

The vector lengths of NO1/NO2 pairs in eight diastereomers with a screw-symmetric H-bond chain have a more or less linear relationship in which the structure with the translational symmetric H-bond chain does not fit (see Fig.

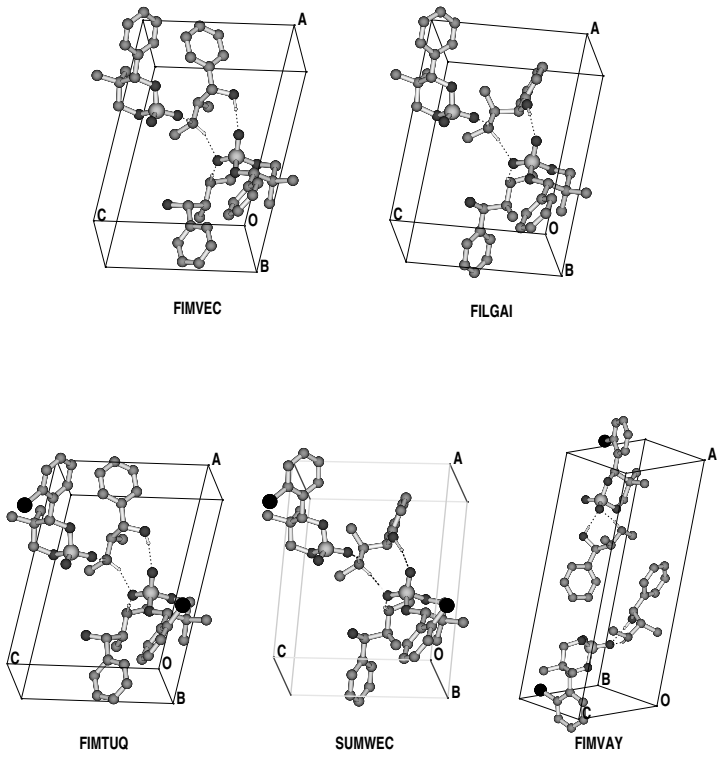


Figure 4.2: The experimental structures of the diastereomers of CPA-E (FIMVEC,FILGAI) and ClCPA-E (FIMTUQ,SUMWEC,FIMVAY) (corresponding CSD reference codes in parenthesis).

4.9).

Figure 4.10 shows the average lengths of NO1 and NO2 as a function of the number of occurences, N, in diastereomers. The diastereomers with screw symmetric H-bonded chains form a cluster with a mean distance of 2.83 $\pm$ 0.03 Å, which is quite different from the diastereomer displaying translationally symmetric H-bonded chains. Apparently the screw symmetric chain has some compensatory flexibility in the lengths of the NO1 and NO2 vectors: when NO1 shortens,

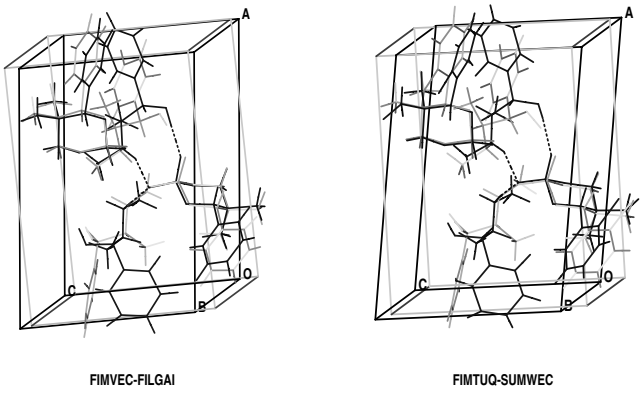


Figure 4.3: Match between the diastereomer pairs of CPA-E (FIMVEC-FILGAI) and diastereomer pairs of ClCPA-E (FIMTUQ-SUMWEC)

NO2 lengthens and vice versa. Some reservations about these conclusions, however, are in order, since the limited number of data points does not allow for truely reliable statistics.

4.3.2 Computational methods

Classical force field methods use Coulomb's law to describe the electrostatic interactions of molecules. This assumes the use of point charges. Most force fields use a quick calculation of point charges based on electronegativity rules .[8,9] Several methods exist to determine the value of local point charges from a quantum mechanical (QM) calculation. These methods can be subdivided into a class where charges are determined by some scheme that partitions the electron density over the atoms (Mulliken,[10] Bader,[11] Hirshfeld[12]) and a class where charges are optimized to reproduce the QM electrostatic potential (ESP) by employing a least-squares fit of the model potential based on point charges to the QM potential. Methods in the latter class differ mainly by how and where the electrostatic potential is sampled.[13,14] These methods sometimes have problems with numerically poorly determined charges on (buried) centres. The restrained ESP method

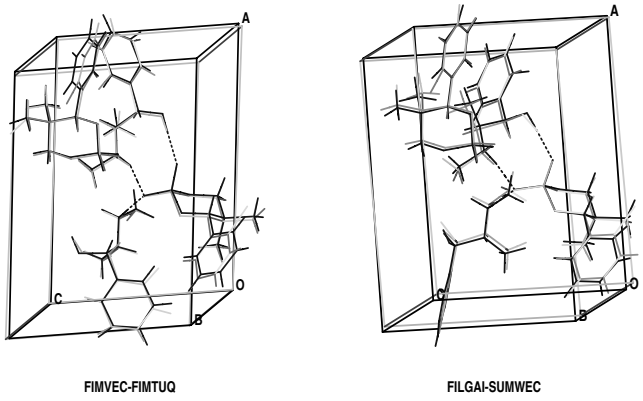


Figure 4.4: Match between the diastereomers of CPA-E: FIMVEC and FILGAI and their ClCPA-E counterparts FIMTUQ and SUMWEC.

(RESP) [15] was developed to deal with this problem by allowing the simultaneous fitting of charges for multiple conformations. In our work we tried firstly to improve the description of the electrostatic interactions by using ESP charges fit to the quantum mechanical electrostatic potential of the anion-cation complex of the diastereomeric salts instead of the ions alone (as done by Leusen). In this way one might expect to incorporate some polarization of the anion caused by the cation and vice versa. We also tried to improve the electrostatics using the RESP method, and finally, we tried to improve the quality of the charges by better sampling of the quantum mechanical electrostatic potential. We accomplished this by sampling the potential on isodensity surfaces as opposed to sampling on a van der Waals surface. We have already developed this method for the same purpose.[16] In combination with these charges, we use the DREIDING 2.21 force field[17] as implemented in the Cerius[2] program[18] for energy minimizations of the cell parameters and the atomic positions.

However, using point charges to describe electrostatic contributions neglects the fact that an atom in the field of other atoms is polarized and produces an electric force which is not spherically symmetric. By representing the molecular

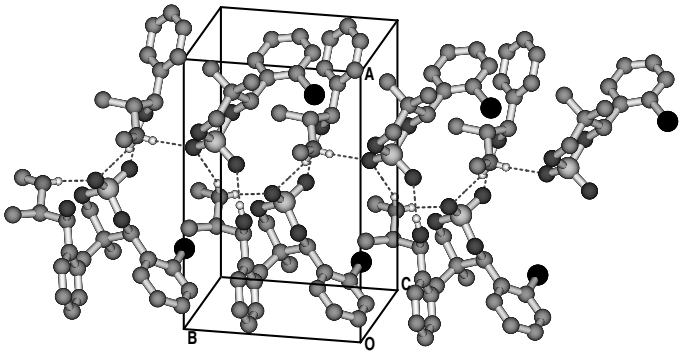


Figure 4.5: Example of the crystal packing class with screw-symmetric hydrogen bonded chains along the b-axis. Three unit cells are shown along this axis

charge distribution as a set of multipoles on a number of centres, the electrostatic interaction can be modelled far more accurately. Stone[19,20] described a generally applicable method for the determination of distributed multipole moments from a quantum mechanically determined electron density. However, most force field methods cannot handle multipoles and rely on partial charges to describe electrostatic interactions. Exceptions are the TINKER package by Ponder,[21] an adaptation of the TINKER package by Mooij et al.[22,23] incorporating polarizable multipoles and the DMAREL program .[24,25] Our computational chemistry program MOLDEN[26] was adapted to calculate distributed multipole moments from a quantum mechanical wavefunction and to interface with both the Mooij version of TINKER and the DMAREL program. Only intermolecular interactions can be optimized with DMAREL: the intramolecular conformations remain rigid. We used the DMAREL program to perform force field minimizations of the cell parameters and the rigid bodies of the ephedrine and cyclic phosphoric ions.

The methods described above all use the properties of the isolated ions or complexes to describe the periodic system. In a more complete description of a periodic system the ions/complexes feel the presence of their symmetry and translationally related copies, resulting in polarization and or charge transfer

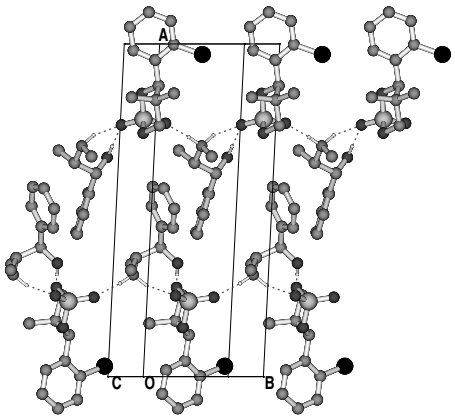


Figure 4.6: Example of the crystal packing class with translationally-symmetry hydrogen bonded chains along the b-axis. Three unit cells are shown along this axis

effects. Quantum mechanical packages that can handle periodic systems incorporate these effects. These packages usually employ density functional theory (DFT)[27,28] for the energy evaluation. Unlike the Hartree-Fock (HF) method, that completely neglects electron correlation, within DFT it can be taken into account to a large extent. DFT is computationally less demanding than other methods for incorporating electron correlation, such as the post-HF singles and doubles configuration interaction (SDCI)[29] and second order Møller-Plesset perturbation theory (MP2).[30]

Table 4.2 shows a summary of the strengths and weaknesses of the density functional theory (DFT) in terms of the different types of interaction present in the system studied. Hydrogen bonding is well described by DFT when used within the generalized gradient approximation (GGA).[31,32] The lack of long range dispersion in exchange-correlation functionals within the generalized gradient approximation, yields a purely repulsive van der Waals potential.[33,34] Meijer[35] and Hobza[32] reported that frequently used exchange-correlation functionals yield a purely repulsive benzene dimer potential. From these considerations it can be

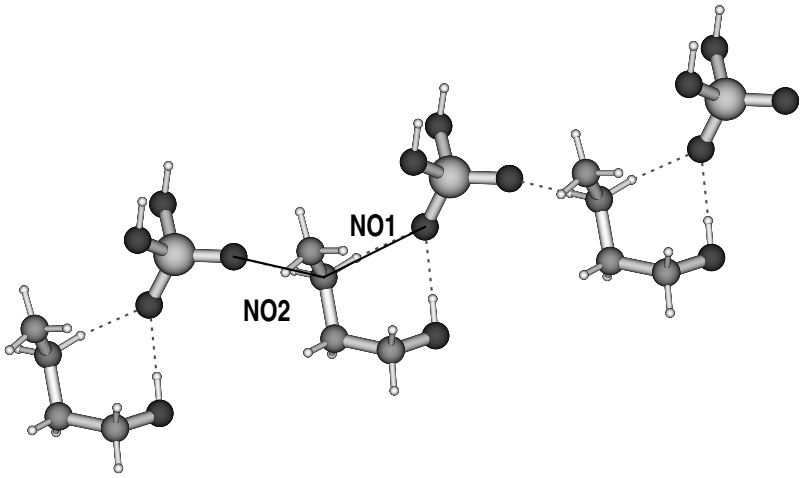


Figure 4.7: Small model translationally symmetric H-bonded chain used for geometry optimizations

concluded that differences in lattice energy between diastereomeric salts can only be calculated accurately with DFT if the contributions to the total energy from ring-ring and van der Waals interactions in the different diastereomers are very similar. Recently, Rovira[36] corrected for the lack of long range dispersion by adding a empirical correction term to the total energy. This however, requires extensive calibration of the correction term with MP2 calculations.

We used the density functional packages VASP, SIESTA and DMol^3 to optimize the atomic positions in the cell, keeping the cell parameters fixed at the experimental values. VASP is the *ab initio* total energy and molecular-dynamics program (Vienna Ab Initio Simulation Program) developed at the Institut für Theoretische Physik of the Technische Universität Wien.[37–39] VASP is a density functional code using a plane wave basis set. The interaction between ion-cores and electrons is described using ultrasoft Vanderbilt pseudopotentials (US-PP),[40,41] so only the valence electrons are treated explicitly. VASP allows the

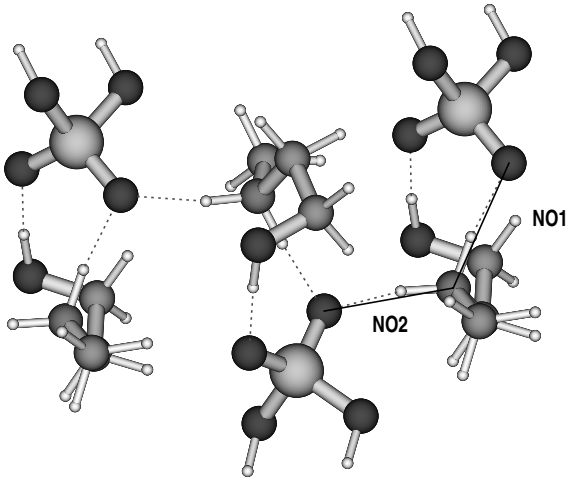


Figure 4.8: Small model screw-symmetric H-bonded chain used for geometry optimizations

simultaneous optimization of atomic positions and cell parameters. The DMol^3[42] package uses atom-centred numerical basis functions and a numerical integration scheme for matrix elements. The numerical integration has the property that the computational effort can go down with decreasing overlap of the basis functions. This property allows us to perform all-electron calculations on our 116 atom systems (the asymmetric unit contains only 58 atoms, but DMol^3 cannot handle point-group symmetry). No core pseudo-potentials were used, since DMol^3 only provides them for elements Sc through Lr. DMol^3 does not allow the optimization of cell parameters.

The SIESTA[43,44] package also uses atomic numerical basis functions and numerical integration. Exploiting the spatial localization of the atomic basis set, the computational cost necessary for the calculation and storage of the self-consistent Hamiltonian matrix is made to scale linearly ($O(N)$) with the number of electrons N in the unit cell. Furthermore, for insulators the evaluation of the density and total energy can be also efficiently performed in $O(N)$ operations using the

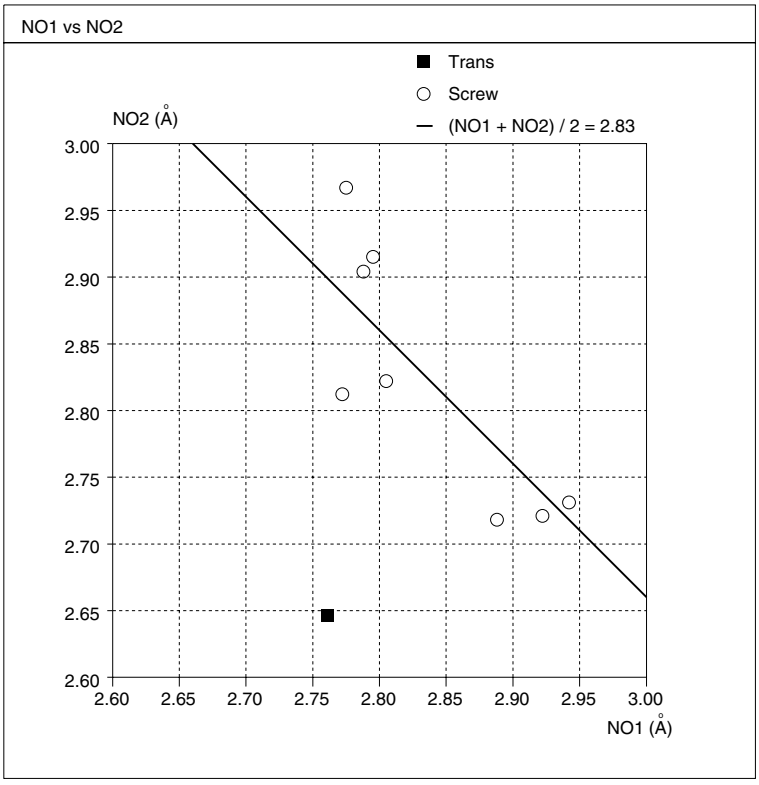


Figure 4.9: Relation between the lengths of the experimental vectors NO1 and NO2 for a number of diastereomers with H-bonded chains with screw and translational symmetry. See Figures 4.8 and 4.7 for the definitions of NO1 and NO2

recently developed $O(N)$ techniques.[45,46] In the present work however, due to the need of computing highly accurate total energies, we have decided to use a standard diagonalization of the Hamiltonian. The interaction between ion-cores and electrons is described using the norm conserving Troullier-Martins pseudopotentials.[47]

The PW91[48] GGA exchange- correlation functional was used with both DMol^3 and VASP. The PBE[49] GGA exchange- correlation functional was used with SIESTA.

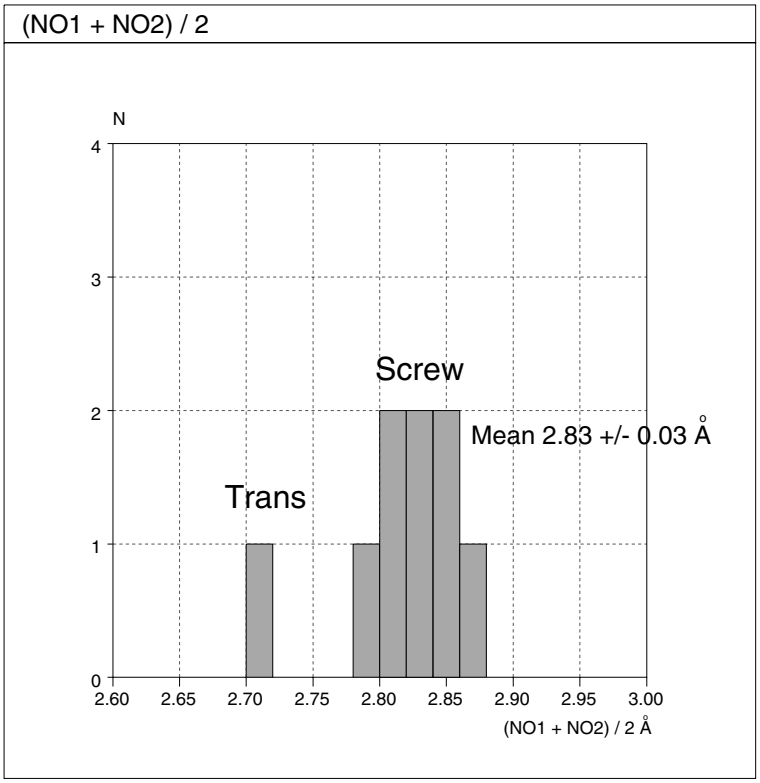


Figure 4.10: The average of distances NO1 and NO2 as a function of the number of occurrences N in diastereomers of derivatives of CPA-E (nine in total). The diastereomers with H-bonded chains with screw and translational symmetry form separate clusters. See Figures 4.8 and 4.7 for the definitions of NO1 and NO2

Another important aspect to consider when performing quantum mechanical calculations on periodic systems is the sampling of the k-space. DMol^3 only considers one special point in the k-space [c], the Γ point. With the VASP and SIESTA packages it is possible to sample the k-space over more points than just Γ, employing the method developed by Monkhorst and Pack.[50] We use the Γ

[c] This is the case for the DMol^3 version 3.9 supplied to us under courtesy license from MSI. The DMol^3 versions 4.2 and higher do allow for k-space integration.

Table 4.2: Strengths and weaknesses of the density functional method

Interaction	*Included*	*Comment*
Ionic complexes	+	
Hydrogen bonds	+	provided non-local DFT is used
Phenyl-phenyl ring	−	See reference[35]
van der Waals	−	Purely repulsive van der Waals potential

k-point for geometry optimization with VASP and investigated the dependence of the relative lattice energies of the diastereomeric salts by performing single point energy calculations with $2 \times 2 \times 2$ Monkhorst-Pack sampling of the k-space on Γ optimized geometries.

Computational resources for the SIESTA calculations allowed us to perform geometry optimizations with fixed cell dimensions at various levels of k-space sampling, as well as to investigate the influence of the simultaneous optimization of cell parameters and atomic positions on the relative energies of the ClCPA-E diastereomers.

All the above-mentioned packages use Ewald summation[51] or related schemes[43] to do the lattice sums of the long-range electrostatic interactions.

4.4 Results

4.4.1 Ab initio calculations on model systems

The next section will deal with the construction of the models in two steps. It will be followed by a section discussing the results of the calculations on the models.

Model construction

Step 1

As explained above, the most important interaction in these diastereomers is in the hydrogen-bridged chains running along the crystallographic b-axis. These chains consist of ionic complex units of protonated ephedrine and the anion of cyclic phosphoric acid. Two types of chains are found, one with screw symmetry and the other with translational symmetry only. Model structures three complexes units long were constructed for both screw symmetric and translationally symmetric forms (see Figures 4.8 and 4.7 respectively). These models were optimized, keeping their respective symmetries, at both Hartree-Fock and density functional level of theory.

The screw and translational symmetries were imposed using a special construction of internal coordinates within the Z-matrix description of the molecular geometry. Table 4.3 shows this Z-matrix construction. Displayed in normal font are the atom labels with a numerical subscript denoting the number of the complex unit they belong to. The first column defines the atom. The second, fourth and sixth columns specify the atomic connectivity with respect to previously defined atoms. The label X denotes a geometrical "auxiliary atom" (or dummy atom). Displayed in bold are the internal variables, some of which are kept constant (numeric values and the string CONST). The third, fifth and seventh columns hold the internal variables of type: bond distance, bond angle and dihedral angle respectively. The automatic generation of these Z-matrix constructs was built into our modelling program MOLDEN.[26]

The implementation of symmetry in this way is only an approximation since:

- The three complexes units in the chains have different environments. The system in this model is therefore not truly periodic, but only quasi-periodic.
- In infinitely long chains, each complex unit would have nearest-neighbor interactions with the complex units before and after it in the chain. In the three-unit chain this is only true for complex unit 2. Complex unit 1 interacts only with the complex unit after it (2) and complex unit 3 interacts only with the complex before it. So in the three unit long chains, there are only two intercomplex interactions (1 with 2 and 2 with 3), whereas in

Table 4.3: The Z-matrix construction to implement 2-fold screw and translational symmetry

X_0						
X_1	X_0	**xx**				
A_1	X_1	**dx**	X_0	**90**		
B_1	A_1	**d1**	X_1	**angx1**	X_0	**dihx1**
C_1	B_1	**d2**	A_1	**ang1**	X_1	**dihx2**
...						
X_2	X_1	**xx**	A_1	**90**	X_0	**180**
A_2	X_2	**dx**	X_1	**90**	A_1	**CONST**
B_2	A_2	**d1**	X_2	**angx1**	X_1	**dihx1**
C_2	B_2	**d2**	A_2	**ang1**	X_2	**dihx2**
...						
X_n	X_{n-1}	**xx**	A_{n-1}	**90**	X_{n-2}	**180**
A_n	X_n	**dx**	X_{n-1}	**90**	A_{n-1}	**CONST**
B_n	A_n	**d1**	X_n	**angx1**	X_{n-1}	**dihx1**
C_n	B_n	**d2**	A_n	**ang1**	X_n	**dihx2**

- For screw symmetry **CONST** = **180**°
- For translational symmetry **CONST** = **0**°, and **xx** is a constant.

infinitely long chains there would be three intercomplex interactions for every three complex units.

- Since the model system is only three units long, the lattice sums over the electrostatic and other long range interactions between one complex unit and all of its symmetry-related copies are not complete.

Step 2

After optimization, these *small model chains*, as we call them from here on, were expanded with cyclic phosphoric acid rings and methyl groups at the ephedrine-like part (see Figures 4.11 and 4.12). We will refer to them as *expanded model*

chains. The corresponding complex units will be referred to as *model complexes* from here on. The full optimization of the expanded model chains was not feasible due to limited computer resources. Therefore we constructed the expanded model chains from the optimized small model chains and model complexes.

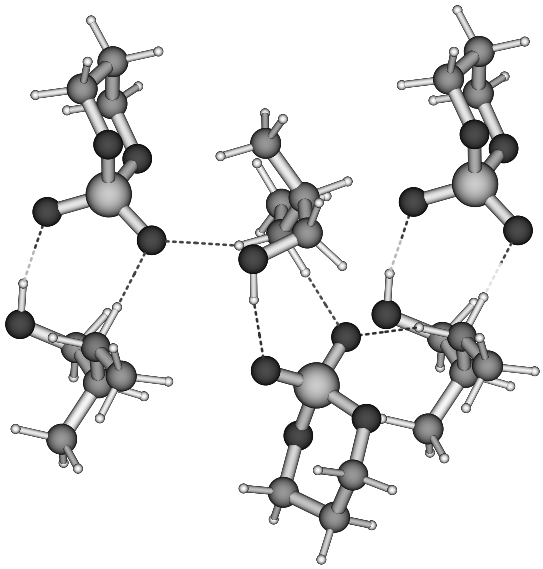


Figure 4.11: Expanded model screw-symmetric H-bonded chain used for single point energy calculations

The model complexes were fully optimized, at both Hartree-Fock and density functional level of theory (see Figure 4.13). The values of internal variables thus obtained for the cyclic phosphoric ring and the ephedrine methyl groups were used in the construction of the *expanded model chains*. Finally, single point energy calculations were performed for the expanded model chains, at the Hartree-Fock, density functional and MP2 level of theory. (MP2 only at the expanded version of the Hartree-Fock optimized small model chain geometry).

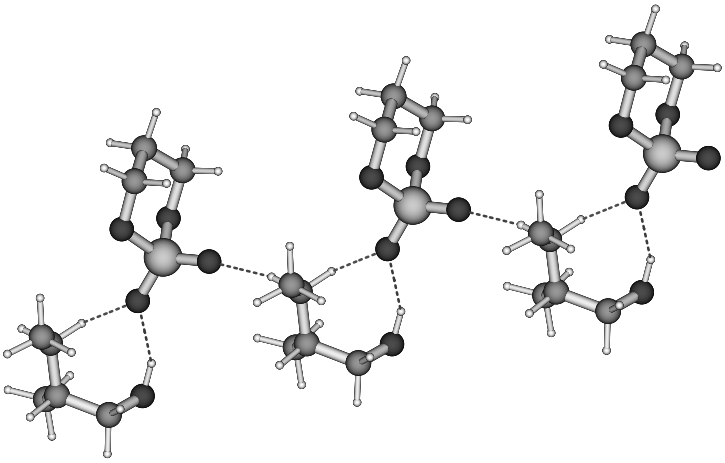


Figure 4.12: Expanded model translationally symmetric H-bonded chain used single point energy calculations

Model calculations

The influence of the basis set superposition error (BSSE) on the relative energies of the model complexes with respect to its constituent ions was investigated using the counterpoise method.[52] All density functional calculations were done using the B3LYP exchange-correlation functional[53] within the program Gaussian94.[54] The B3LYP functional has been reported to give good energetic results for hydrogen-bonded systems,[31,32] even in combination with a medium sized basis-set such as 6-31+G(d,p).[55] A customized basis-set derived from the 6-31+G(d,p) basis-set was used for geometry optimizations:

- The use of diffuse basis functions on all atoms leads to linear dependencies, so they were used only on the oxygen atoms. Without diffuse basis functions on the oxygens, one of the hydrogens attached to the ephedrine nitrogen migrates back to cyclic phosphoric acid oxygen during geometry optimization.

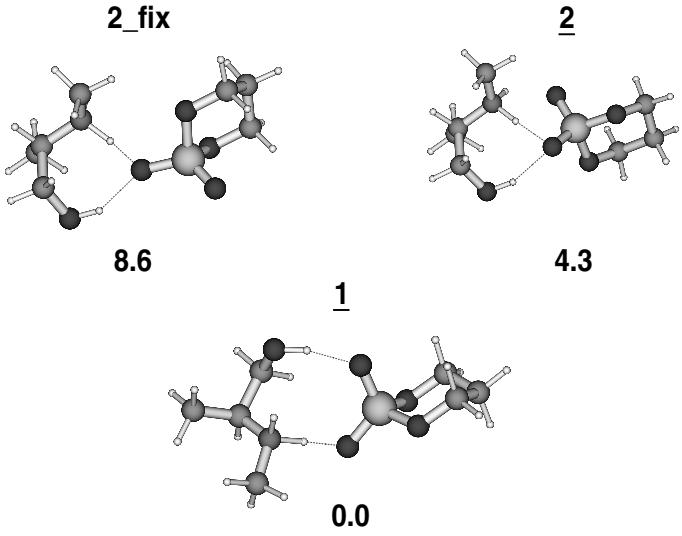


Figure 4.13: Model complexes optimized at the density functional level of theory with the B3LYP functional and a customized basis-set, with their relative energies at the same level of theory, corrected for the BSSE

- In addition, the hydrogen polarization functions were used only for those hydrogen atoms that are involved in hydrogen-bonding.

Figure 4.13 shows the fully optimized geometries of the model complexes at the B3LYP density functional level together with their relative energies corrected for BSSE. The model complex corresponding to the screw-symmetric hydrogen-bonded chain (labelled 1) was found to be 4.3 kcal/mol more stable than its translationally symmetric chain counterpart (labelled 2). Figure 4.14 shows the remarkable agreement between the optimized structure of model complex 1 (grey) and the corresponding part of the experimental structure FILGAI (black). The fully optimized structure of model complex 2 does not match the experimental structure very well. In the experimental structure the orientation of the cyclic phosphoric ring is such that one of the oxygens connected to the phosphorus atom

can interact with the next complex in the hydrogen-bonded chain.

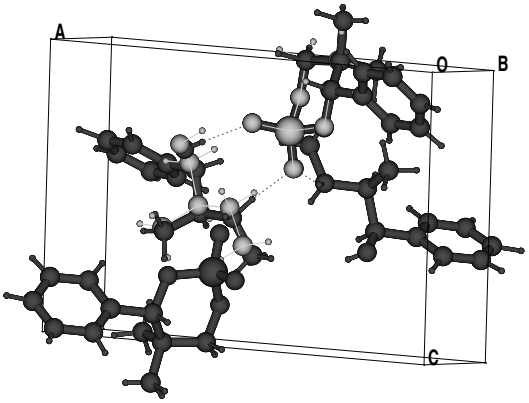


Figure 4.14: Match between B3LYP optimized model complex1 (grey) and experimental structure FILGAI

Figure 4.15 shows the improved match between model complex 2 (grey) optimized while restraining the torsional angle that determines the orientation of the cyclic phosphoric ring to the experimental value, with the experimental structure FIMVAY (black). The partially fixed model complex 2 (labelled 2_fix in figure 4.13) is 8.6 kcal/mol less stable than model complex 1. Model complex 1 and the partially fixed model complex 2 can be considered as the geometrical building blocks of hydrogen-bonded chains with screw (chain1) and translational (chain2) symmetry respectively.

In tables 4.4 and 4.5 the absolute and relative energies at the Hartree-Fock, B3LYP and MP2 level of theory are given for the small model chains (chain1 and chain2), the expanded model chains and the model complexes 1 and 2. For the small model chains and the model complexes, the fully optimized geometries at the Hartree-Fock level of theory were used, employing the customized basis-set described above. For the calculation of the single point energies the 6-31+G(d,p) basis-set was used. The energy difference between the model complex 1 and partially fixed model complex 2 (9.9, 8.6 and 9.2 kcal/mol for HF, B3LYP and

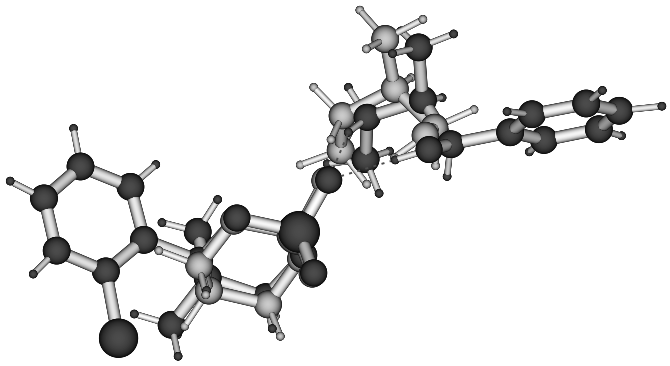


Figure 4.15: Match between B3LYP optimized model complex1 (grey) and experimental structure FIMVAY

MP2) is largely conserved between the two chains. The energy differences between the small model chains are 8.7, 7.8 and 11.5 kcal/mol per complex for HF, B3LYP and MP2 respectively. The energy differences between the expanded model chains are 7.3 and 6.1 kcal/mol per complex for HF and B3LYP respectively. Hartree-Fock and B3LYP calculations agree within 1.2 kcal/mol.

In Table 4.6 the absolute and relative energies at the B3LYP level of theory are given for the fully optimized small model chains (chain1 and chain2), the expanded model chains and the fully optimized model complexes 1 and 2. All optimizations were performed at the B3LYP level of theory, using the customized basis-set described above. For the single point energies the 6-31+G(d,p) basis-set was used. The BSSE corrected relative energies are included for the model complexes. The largest correction was 0.2 kcal/mol, which is in the range of the error in the experimental lattice energies. The energy difference between the two unit complexes (8.6 kcal/mol) is largely conserved between the two small model chains (24.1 kcal/mol for 3 units; 8.0 kcal/mol per unit).

Figure 4.16 shows the excellent match of the *ab initio* (B3LYP) translational type H-bonded chain with the diasteromer of the chlorine derivative of CPA with

*Table 4.4: Absolute energies (in Hartrees) at different levels of theory employing the 6-31+G** basis-set for model complexes and model chains, optimized at the Hartree-Fock level of theory using a customized basis-set*

compound	E_{HF}	E_{B3LYP}	E_{MP2}
$complex1$	−1045.193870	−1049.954339	−1047.423529
$complex2$	−1045.188253	−1049.948541	−1047.417045
$complex2_{fixed}$	−1045.178145	−1049.940641	−1047.408882
$chain1$	−2670.727012	−2681.784810	−2675.752116
$chain2$	−2670.685494	−2681.746754	−2675.697005
$chain1_{expanded}$	−3135.624808	−3149.898913	–
$chain2_{expanded}$	−3135.589943	−3149.869906	–

ephedrine (CSD reference code FIMVAY). Figure 4.17 shows the match of the *ab initio* (B3LYP) screw type H-bonded chain (grey) with the diasteromer of doubly substituted chlorine derivative of CPA with ephedrine (black), with CSD reference code KOSYIA. The *ab initio* geometry displayed is of an intermediate structure 1.1 kcal/mol less stable than the final optimized structure. The match of the final optimized structure with KOSYIA is less good. Probably the omission of methyl-groups on the ephedrine-like part of the chain during the optimization allows low energy geometries that would no longer be low energy geometries if the structure were to be expanded with these methyl-groups. This is confirmed by the relative energies of the fully expanded model chains in table 4.6 (denoted by ++). The screw type chain ($chain1^{++}_{expanded}$) is only 1.1 kcal/mol more stable than the translational type chain ($chain2^{++}_{expanded}$), versus 24.1 kcal/mol in the unexpanded case. In $chain1^{++}_{expanded}$ there is a close contact of 1.95 Å between a methyl group

*Table 4.5: Relative energies (in kcal/mol) at different levels of theory employing the 6-31+G** basis-set for model complexes and model chains, optimized at the Hartree-Fock level of theory using a customized basis-set*

compound	E_{HF}	E_{B3LYP}	E_{MP2}
$complex1$	0	0	0
$complex2$	3.5	3.6	4.1
$complex2_{fixed}$	9.9	8.6	9.2
$chain1$	0	0	0
$chain2$	26.1	23.9	34.6
$chain1_{expanded}$	0	0	–
$chain2_{expanded}$	21.9	18.2	–

hydrogen and an atom on the cyclic phosphoric ring that is clearly energetically unfavourable. This is also supported by the relative energies of the not fully expanded model chains (denoted by +), where not fully expanded means that only cyclic phosphoric rings are added and not methyl groups. Figure 4.18 compares the calculated NO1 and NO2 distances (as defined in Figures 4.8 and 4.7), with the experimental NO1 and NO2 distances found in eight diastereomers with H-bonded screw-symmetric chains and one with translational symmetry. The calculated NO1/NO2 pair for the H-bonded screw-symmetric chain lies within the same area as the corresponding experimental values. The calculated NO1/NO2 pair for the H-bonded translationally symmetric chain and its experimental counterpart lie well outside this area. The screw type chain ($chain1^{+}_{expanded}$) is 23.4 kcal/mol more stable than the translational type chain ($chain2^{+}_{expanded}$). This matches the small model chains. This shows the limitations of using a model

Table 4.6: Absolute and relative energies (in Hartrees and kcal/mol respectively) at B3LYP level of theory for model complexes and model chains, geometry optimized at the B3LYP level of theory using a customized basis-set

compound	$E_{Abs.}$	$E_{Rel.}$	$E_{Rel.+BSSE}$
$complex1$	−1049.937969	0	0
$complex2$	−1049.931441	4.1	4.3
$complex2_{fixed}$	−1049.924316	8.6	8.6
$chain1$	−2681.735734	0	–
$chain2$	−2681.697401	24.1	–
$chain1^{+}_{expanded}$	−3031.909557	0	–
$chain2^{+}_{expanded}$	−3031.872285	23.4	–
$chain1^{++}_{expanded}$	−3149.815590	0	–
$chain2^{++}_{expanded}$	−3149.813798	1.1	–

$^{+}$ small + cyclic phosphorus rings
$^{++}$ small + cyclic phosphorus rings + methyl

system that is too small to describe a large system.

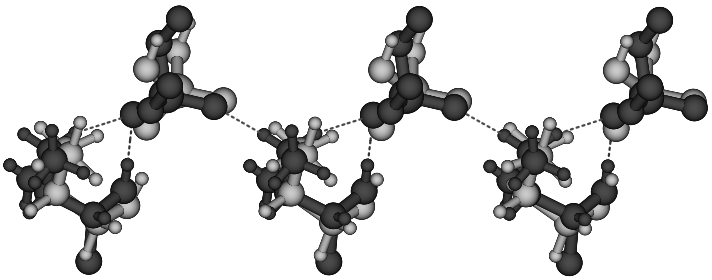


Figure 4.16: Match of the ab initio translationally-symmetric H-bonded chain with one diasteromer of ClCPA-E (CSD refcode FIMVAY)

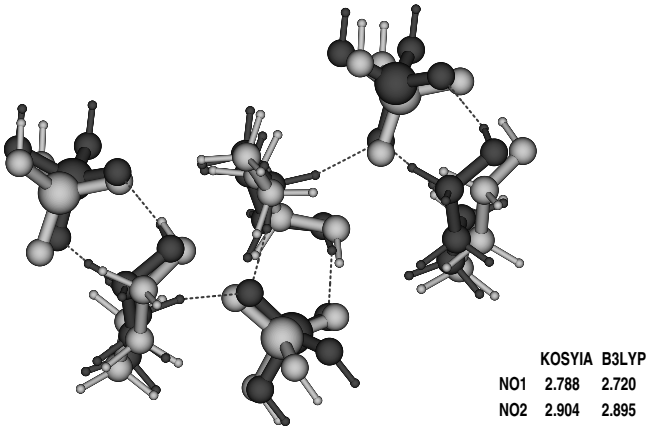


Figure 4.17: Match of the ab initio screw-symmetric H-bonded chain (intermediate geometry, 1.1 kcal/mol away from the optimum) with the diasteromer of doubly substituted chlorine derivate of CPA-E (CSD refcode KOSYIA)

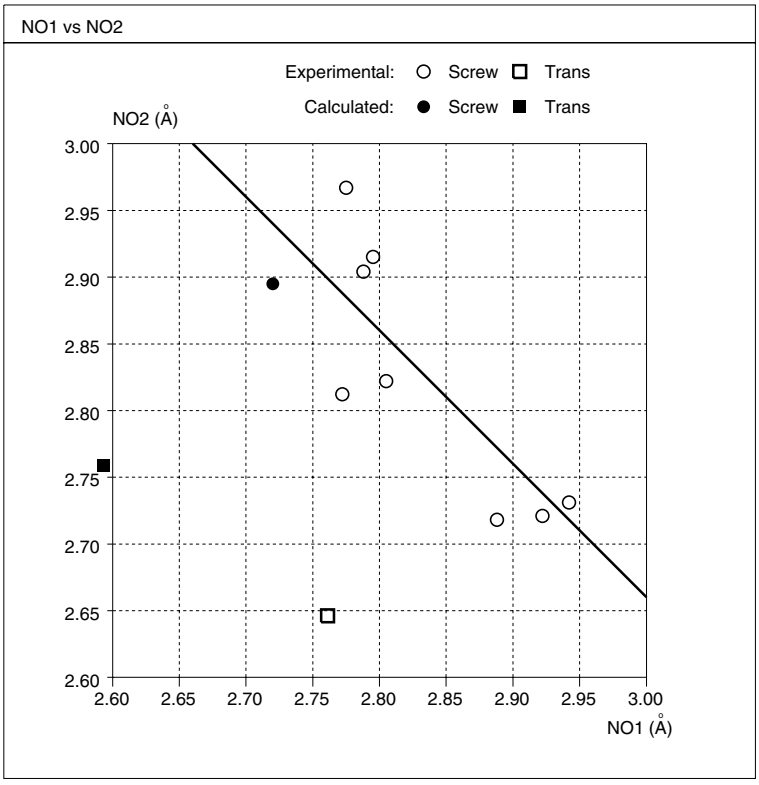


Figure 4.18: The experimental distances NO1 and NO2 for a number of diastereomers with H-bonded chains with screw and translational symmetry, compared with the calculated NO1 and NO2 distances. See figures 4.8 and 4.7 for the definition of NO1 and NO2

In conclusion, one can say that the screw type chain is roughly 8 kcal/mol per complex unit more stable than the translational type chain. The exclusion of certain parts of the large system can cause the optimized structure of the model system to deviate from the optimal structure of the large system, which in turn can cause the expanded model to have an unrealistic geometry.

4.4.2 Lattice energy calculations using a classical force field

The following charge sets were used in combination with the DREIDING force field:

- set1: ESP fit on the whole complex
- set2: ESP fit on the whole complex, combined with the RESP method
- set3: ESP fit on the ions
- set4: ESP fit on the ions, combined with the RESP method
- set5: ESP sampled on isodensity surfaces fit on the ions
- set6: ESP sampled on isodensity surfaces fit on the ions, combined with the RESP method.

The charge sets 1 and 2 were derived from a QM calculation of the partially optimized complexes at the Hartree-Fock level using the customized basis-set discussed earlier, where "partially optimized" means that the internal variables which were thought to be sensitive to the crystal environment were kept fixed, while all others were relaxed. In this way the influence of errors in the experimental crystal structures on the charge distribution is minimized. Several of the crystal structures contained poor hydrogen positions, with, for example, hydrogens connected to a phenyl ring having an H-C-C bond angle of 99°, instead of the common 120°. These optimizations were computationally demanding so we decided not to include the SUMWEC structure in these calculations.

Table 4.7 shows the total charge on the CPA substructure in the diastereomers of CPA-E and ClCPA-E, for charge sets 1 and 2. The total electronic charge on CPA is roughly -0.8, which is significantly less than -1.0, the value for an isolated

Table 4.7: Total charge on the CPA substructure in the diastereomers of CPA-E and ClCPA-E, for charge sets 1 and 2.

Refcode	*set*1	*set*2
FIMVEC	−0.84	−0.84
FILGAI	−0.84	−0.84
FIMTUQ	−0.83	−0.76
FIMVAY	−0.78	−0.76

CPA anion. This charge transfer effect can be seen as a depolarization of the complex with respect to the isolated ions.

Table 4.8 shows the absolute and relative lattice energies and percentage errors in lattice parameters of the diastereomers of CPA-E and ClCPA-E optimized with the DREIDING-2.21 force field in combination with charge set 1. Ewald summation was used for the non-bonded interactions. The default settings were used. The inter- and intra- molecular geometry are optimized simultaneously. The structural agreement between calculated and experimental structures is relatively good, with errors on average around 2 percent. The only unfavourable exception is the b parameter of the FIMVEC structure, which is around 8 percent in error. The agreement between the calculated and experimental lattice energies is, however, very poor, both in the magnitude of the energy differences between the diastereomeric salt pairs FIMVEC/FILGAI and FIMTUQ/FIMVAY and the relative stability order of the latter. Experiment shows the FIMVAY structure to be 2.3 kcal/mol more stable than the FIMTUQ structure, while DREIDING calculates it to be 11.0 kcal/mol less stable.

As shown in Table 4.9, charge set 2 gives an even better structural agreement with the experiment than charge set 1. In addition, the magnitude of the energy differences between the diastereomeric salt pairs FIMVEC/FILGAI and

Table 4.8: Absolute and relative lattice energies and percentage errors in lattice parameters of the diastereomers of CPA-E and ClCPA-E, calculated with the DREIDING force field using charge set 1

Refcode	Δa (%)	Δb (%)	Δc (%)	$\Delta \beta$ (%)	E_{abs} (kcal/mol)	E_{rel} (kcal/mol)
FIMVEC	−2.5	7.7	−1.9	2.2	8.1	0
FILGAI	1.2	1.1	0.6	−2.6	13.8	5.6
FIMTUQ	2.1	−0.5	0.5	−1.3	1.8	0
FIMVAY	0.9	1.0	2.3	−0.9	12.8	11.0

FIMTUQ/FIMVAY are more in agreement with experiment. The relative stability order of the FIMVAY and FIMTUQ structures is still predicted wrongly, however.

All other charge sets also give good structural agreement between calculation and experiment. Table 4.10 lists the relative lattice energies for the diastereomers of CPA-E and ClCPA-E, for all the six charge sets. All charge sets give an incorrect order of stability for the salt pair FIMTUQ/FIMVAY. The predicted energy difference ranges from 3 to 19 kcal/mol. Clearly the DREIDING force field is extremely susceptible to variations in the point charges. The reason for this can be found in Table 4.11, which lists the intermolecular part of the relative lattice energies for the various charge sets. The intermolecular part of the relative lattice energies for the salt pair FIMTUQ/FIMVAY, with the exception of charge set 1, is much less variable than the total relative lattice energies and ranges from 5 to 8 kcal/mol. Charge set 1 behaves differently, because it is the only charge set for which the total charge on the ions is different for FIMTUQ and FIMVAY (0.83 versus 0.78). Obviously this charge transfer effect has a large effect on the intermolecular interaction of the anion and cation. The large variation in the total

Table 4.9: Absolute and relative lattice energies and percentage errors in lattice parameters of the diastereomers of CPA-E and ClCPA-E, calculated with the DREIDING force field using charge set 2

Refcode	Δa (%)	Δb (%)	Δc (%)	$\Delta \beta$ (%)	E_{abs} (kcal/mol)	E_{rel} (kcal/mol)
FIMVEC	0.5	3.3	−0.6	0.2	15.9	0
FILGAI	1.4	0.7	0.6	−3.0	16.0	0.2
FIMTUQ	2.0	−0.2	0.8	−0.9	2.0	0
FIMVAY	0.9	1.2	2.2	0.0	5.3	3.3

relative lattice energies must be attributed to variations in the *intra*molecular contributions to the total relative lattice energies. The DREIDING force field includes intramolecular charge-charge interactions for atoms which are more than two bonds away from each other. Since ESP charges are designed to reproduce the electrostatic potential outside the molecule, there is no reason to assume they should be optimal for reproducing intramolecular energies. An inspection of the individual charges in all six charge sets shows that some charges associated with atoms not very close to the molecular surface show large variations between the different sets (variations of 0.4, sometimes accompanied by sign reversal). This is probably because these charges contribute little to the actual value of the charges-derived electrostatic potential at the molecular surface particularly in the case of the ions, where the potential is dominated by a few strong charge centres. Although these large charge variations may have little effect on the potential at the molecular surface, they may have a large effect when used to calculate intramolecular charge interactions.

In conclusion, one can say that only fitting charges on the complexes rather than the ions, in combination with fitting on multiple conformations (RESP),

Table 4.10: Relative lattice energies in kcal/mol for the diastereomers of CPA-E and ClCPA-E, calculated with the DREIDING force field using various charge sets

	$set1$	$set2$	$set3$	$set4$	$set5$	$set6$	exp.
$E_{FILGAI-FIMVEC}$	5.6	0.1	5.1	1.1	−4.3	1.4	-0.1
$E_{FIMVAY-FIMTUQ}$	11.0	3.3	16.7	11.9	18.9	9.1	-2.8

Table 4.11: Intermolecular part of the relative lattice energies in kcal/mol for the diastereomers of CPA-E and ClCPA-E, calculated with the DREIDING force field using various charge sets.

	$set1$	$set2$	$set3$	$set4$	$set5$	$set6$
$E_{FILGAI-FIMVEC}$	2.2	4.0	3.3	4.1	4.0	4.8
$E_{FIMVAY-FIMTUQ}$	20.1	5.1	7.8	7.8	6.5	6.4

gives lattice energy differences between diastereomeric salt pairs of the correct magnitude, but the relative stability order is still incorrect. The use of isodensity sampling of the electrostatic potential does not yield a systematic improvement over van der Waals surface sampling for this system.

4.4.3 Lattice energy calculations using distributed multipole expansions

The DMAREL package [25] incorporates electrostatic contributions to the interaction energy via interacting multipole expansions. Only intermolecular interactions can be optimized with this method: the intramolecular conformations remain rigid. Calculations with the DMAREL package were done for our model system, the diastereomeric salts of CPA-E and ClCPA-E, choosing the cyclic phosphoric and ephedrine parts as rigid bodies. The multipole expansions were externally generated via our MOLDEN package.[26] The repulsion-dispersion parameters for hydrogen, carbon, nitrogen and oxygen provided with the program were used (the parameter set named FIT+0.9DMA in reference[24]). The program provides no such parameters for chlorine and phosphorus atoms. For chlorine, the parameters were taken from Williams *et al.*[56] As with the parameters for hydrogen, carbon, nitrogen and oxygen, these parameters were optimized in the presence of atomic charges in the nonbonded potential. For phosphorus no such parameters were available. In our model system phosphorus does not have close intermolecular contacts, since it is tetrahedrally surrounded by oxygens. The exact values of the phosphorus parameters therefore will have little effect on the relative energies of the diastereomers. We chose the chlorine parameters to describe the repulsion-dispersion interactions of phosphorus.

Cox and Williams[57] found that most Hartree-Fock wavefunctions with split-valence basis sets overestimate molecular dipole moments. They introduced a dipole scaling factor of 0.9 which is widely used in molecular modeling. The authors of DMAREL concluded that scaling of the multipoles with a factor of 0.9 provides the best predictions of the lattice energies of the non-hydrogen-bonded molecules, but generally underestimates the lattice energies of hydrogen-bonded molecules.[24] Since our model system contains both hydrogen-bonds and ionic interactions, we also investigated the influence of the multipole scaling factor on the lattice energies.

Tables 4.12, 4.13 and 4.14 show the absolute and relative lattice energies and percentage errors in lattice parameters of the diastereomers of CPA-E and ClCPA-E, calculated with the DMAREL program using factors of 0.7, 0.9 and 1.0 respectively to scale the multipole moments.

Table 4.12: Absolute and relative lattice energies and percentage errors in lattice parameters of the diastereomers of CPA-E and ClCPA-E, calculated with the DMAREL program using a factor of 0.7 to scale the multipole moments

Refcode	Δa (%)	Δb (%)	Δc (%)	$\Delta\beta$ (%)	E_{latt} (kJ/mol)	E_{rel} (kcal/mol)
FIMVEC	1.9	1.8	−3.9	−2.4	−277.8	0
FILGAI	2.2	2.2	−3.3	−1.7	−259.5	4.4
FIMTUQ	2.0	−0.4	−2.2	0.03	−275.7	0
SUMWEC	1.2	1.0	−0.7	−0.6	−265.5	2.4
FIMVAY	3.1	4.3	−3.7	4.3	−271.3	1.0

Table 4.13: Absolute and relative lattice energies and percentage errors in lattice parameters of the diastereomers of CPA-E and ClCPA-E, calculated with the DMAREL program using a factor of 0.9 to scale the multipole moments

Refcode	Δa (%)	Δb (%)	Δc (%)	$\Delta \beta$ (%)	E_{latt} (kJ/mol)	E_{rel} (kcal/mol)
FIMVEC	3.8	−1.6	−3.7	−1.5	−372.1	0
FILGAI	2.2	1.5	−4.0	−0.5	−343.4	6.9
FIMTUQ	2.4	−1.3	−2.6	0.9	−356.0	0
SUMWEC	2.0	0.05	−1.5	0.7	−347.4	2.1
FIMVAY	9.5	4.7	−7.7	−1.4	−349.3	1.6

Table 4.14: Absolute and relative lattice energies and percentage errors in lattice parameters of the diastereomers of CPA-E and ClCPA-E, calculated with the DMAREL program using a factor of 1.0 to scale the multipole moments

Refcode	Δa (%)	Δb (%)	Δc (%)	$\Delta\beta$ (%)	E_{latt} (kJ/mol)	E_{rel} (kcal/mol)
FIMVEC	4.4	−2.9	−3.7	−1.1	−431.9	0
FILGAI	2.0	1.2	−4.4	−0.2	−396.1	8.6
FIMTUQ	2.7	−1.6	−3.1	1.3	−405.8	0
SUMWEC	2.7	−0.5	−2.0	1.3	−399.8	1.4
FIMVAY	11.5	6.3	−10.5	−2.3	−419.6	−3.3

The experimental structures are reproduced with an r.m.s. error in the lattice parameters of roughly 2% (as compared to 1% with DREIDING), with the exception of the FIMVAY structure. Especially at the higher scaling constants, large errors up to 12% in individual lattice parameters were found for FIMVAY. Table 4.15 shows the r.m.s. of the percentage error in lattice parameters for each structure as a function of the factor used to scale the multipole moments (ranging from 0.5 to 1.0). The best r.m.s. average over all five stuctures is obtained with a scaling constant of 0.7.

The calculated relative energies are of the correct magnitude (0−3 kcal/mol) for the ClCPA-E diastereomers, but too large for the CPA-E diastereomers (5−9 kcal/mol). The relative stability orders of FIMVAY with respect to FIMTUQ are not correctly predicted except with a multipole scaling constant of 1.0. Here FIMVAY is predicted to be more stable than FIMTUQ ($E^{Calc.}_{FIMVAY-FIMTUQ} = -3.3$ versus $E^{Expt.}_{FIMVAY-FIMTUQ} = -2.8$ kcal/mol).

Lattice energy differences have to be corrected for the energy differences of the rigid bodies of diastereomeric salt pairs, since these are not considered in the DMAREL approach. When treating the constituents of the complexes as

Table 4.15: The root mean square of the percentage error in lattice parameters calculated with the DMAREL program for the diastereomers of CPA-E and ClCPA-E, as a function of the factor used to scale the multipole moments. In parentheses the average is given for when the diastereomer FIMVAY is excluded.

Refcode *scale*	0.5	0.7	0.8	0.9	1.0
FIMVEC	3.4	2.6	2.6	2.9	3.2
FILGAI	3.0	2.4	2.4	2.4	2.5
FIMTUQ	1.3	1.5	1.7	1.9	2.3
SUMWEC	1.7	0.9	1.0	1.3	1.8
FIMVAY	2.7	3.9	4.7	6.6	8.5
average	2.5(2.5)	2.0(2.0)	2.8(2.0)	3.6(2.2)	4.4(2.5)

rigid bodies, only their conformational energy differences need to be considered. Table 4.16 shows the relative energies calculated with the DMAREL program for the diastereomers of CPA-E and ClCPA-E corrected for conformational energy differences, as a function of the factor used to scale the multipole moments. The conformational energy differences between the complex parts (see Table 4.17) were calculated at the density functional level of theory with the DMol^3 program with geometries resulting from periodic ion-relaxation calculations from the VASP program [d]. There is a remarkable agreement between the results with scaling factor 1.0 and the results discussed below for DMol^3 lattice energy calculations (see Table 4.22). The conformational energy corrections dominate the

[d] We could use neither the DMol^3 nor the VASP program for both the energy calculations and geometry relaxations with periodic boundary conditions. Because of limited computer resources, the structures of FILGAI and FIMVEC could not be optimized with the DMol^3 program. The VASP program, on the other hand, has no provisions for non-periodic calculations.

Table 4.16: Relative energies in kcal/mol calculated with the DMAREL program for the diastereomers of CPA-E and ClCPA-E corrected for conformational energy differences, as a function of the factor used to scale the multipole moments

Refcode	*scale*	1.0	0.9	0.7	E_{exp}
FIMVEC		0	0	0	0
FILGAI		5.4	3.7	1.2	−0.1
FIMTUQ		0	0	0	0
SUMWEC		−3.0	−2.3	−2.0	0
FIMVAY		−1.1	3.9	3.3	−2.8

lattice energy differences between the diastereomers. In the single case of the conformational energy correction of SUMWEC with respect to FIMTUQ, a correction of up to 4.4 kcal/mol was found. The relative lattice energy of SUMWEC with respect to FIMTUQ calculated with the DMol^3 program is only half of this: 2.3 kcal/mol, whilst experimentally it is found to be 0.1 kcal/mol. This suggests that the conformational energy corrections calculated with DMol^3 may be insufficiently reliable, at least in some cases. The lattice energy calculations with the DMol^3 and VASP programs are discussed in detail in following sections.

Table 4.17: Conformational energy differences between complex parts of the diastereomers of CPA-E and ClCPA-E, calculated at the density functional level of theory with the DMol^3 program with the geometries resulting from periodic ion-relaxation calculations with the VASP program

Refcode	$E_{ephedrine-part}$	$E_{cyclicphosphoric-part}$	$E_{tot.}$	E_{Rel}
FIMVEC	−520.413021	−1070.077770	−1590.490791	0
FILGAI	−520.418207	−1070.077728	−1590.495935	−3.2
FIMTUQ	−520.412903	−1529.695566	−2050.108469	0
SUMWEC	−520.419698	−1529.695779	−2050.115477	−4.4
FIMVAY	−520.411908	−1529.692974	−2050.104882	2.3

4.4.4 Lattice energy minimizations using VASP

The crystal structures of the diastereomers of CPA-E and ClCPA-E were optimized with VASP, while keeping cell dimensions fixed at the experimental values. The wave functions are expanded in a plane wave basis set. Only waves for which $|k+K|^2 \leq E_{cut}$ are included. An E_{cut} of 37 Ry was required.[58] The pseudopotentials were optimized for use with the PW91 exchange- correlation functional.[48] We took especially care to accurately describe the augmentation charges of the ultra-soft pseudopotentials. Non-linear core corrections[59] were employed for the elements chlorine and phosphorus. The non-local projections were carried out in real-space.[60,61]

The results of these calculations at the Γ-point in k-space are presented in Figure 4.19 and Table 4.18. Figure 4.19 shows the match between VASP-optimized and experimental structures of the diastereomers of CPA-E and ClCPA-E with CSD reference codes: FIMVEC, FILGAI, FIMTUQ, SUMWEC, FIMVAY.

The agreement between calculated and experimental structures is very good, with the exception of the hydrogen positions. This is not surprising since the hydrogen positions determined by X-ray diffraction are always less accurate than those of the heavier elements. Specially in the case of FIMVEC and FIMTUQ where in the experimental structures some hydrogen connected to the ephedrine phenyl ring make CCH angles of 140° and 145°, where they should typically be around 120°. In the case of FIMVAY a small deviation was found in the torsion angle defining the orientation of the phenyl with respect to the cyclic phosphoric ring to which it is attached. Table 4.18 shows absolute and relative energies of the diastereomers of CPA-E and ClCPA-E as calculated with the VASP package. Included for comparison are the experimental energies. *A good agreement between theory and experiment is found.* FIMVAY is predicted to be the most stable chlorine containing diastereomer. Also the energy differences between FIMVEC and FILGAI and its chlorine analogues FIMTUQ and SUMWEC were found to be within the experimental error.

This table also contains a FIMVAY analogue. FIMVAY is the most stable chlorine containing diastereomer. However, its non-chlorine analogue is unknown, suggesting that it is less stable than FIMVEC or FILGAI. The non-chlorine FIMVAY analogue was optimized with VASP, keeping its lattice parameters at the

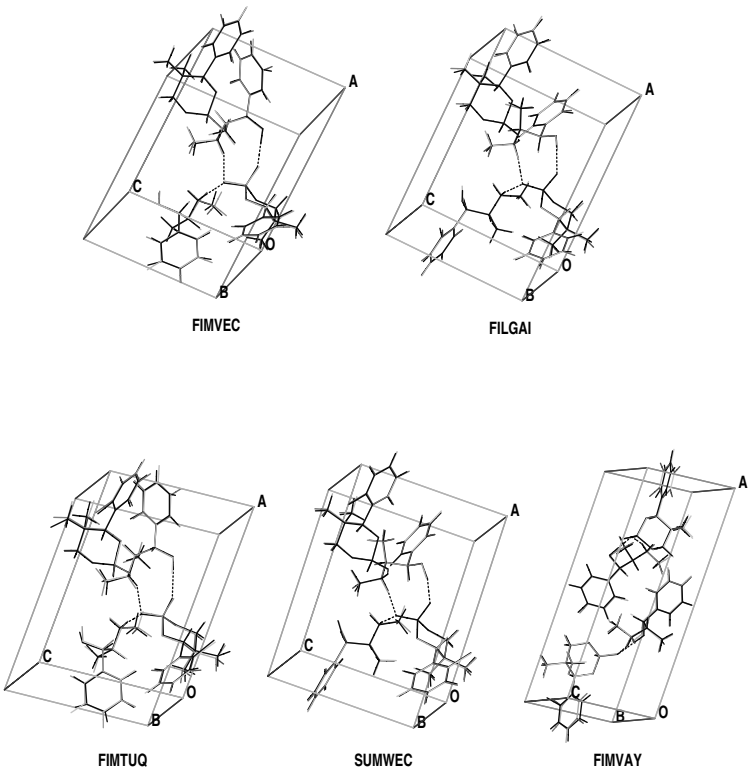


Figure 4.19: Match between VASP optimized (grey) and experimental (black) structures, with CSD refcodes: FIMVEC, FILGAI, FIMTUQ, SUMWEC, FIMVAY

experimental values found for FIMVAY. The calculated structure was indeed found to be less stable than FIMVEC and FILGAI by 1.4 and 2.5 kcal/mol respectively. Figure 4.20 shows the match between the VASP optimized structures of FIMVAY and its non-chlorine analogue. The two structures are very similar, except for the torsion angle that defines the orientation of the phenyl with respect to the cyclic phosphoric ring to which it is attached.

This can be seen even more clearly in Figure 4.21 where the difference in

Table 4.18: Absolute (eV) and relative (kcal/mol) energies of the diastereomers of CPA-E and ClCPA-E, with atomic positions optimized at the density functional level of theory with the VASP program, keeping cell dimensions fixed at experimental values, and only Γ-point was used for the k-space sampling. Experimental energies are also shown for comparison. Included is the non chlorine analogue of FIMVAY, optimized using the FIMVAY cell

Refcode	E_Γ	ΔE_Γ	$\Delta E_{Exp.}$
FIMVEC	−352.743591	0	0
FILGAI	−352.792145	−1.1	−0.1
FIMVAY analogue	−352.682279	1.4	−
FIMTUQ	−351.106934	0	0
SUMWEC	−351.160400	−1.2	0
FIMVAY	−351.207520	−2.3	−2.8

orientations of the phenyl rings is shown between FIMVAY (black) and its non-chlorine analogue (grey). In FIMVAY there is a close contact (2.426 Å) between the chlorine connected to one ring and a hydrogen connected to its translational copy. In its non-chlorine analogue the phenyl rings have turned away, to yield a much less close contact of 3.725 Å between a hydrogen connected to one ring and a hydrogen connected to its translational copy. This clearly suggests that an energetically favourable interaction in FIMVAY is substituted by an energetically unfavourable interaction in its non-chlorine analogue.

Just by looking at the partial (ESP) charges of the FIMVAY chlorine and the hydrogen with which it is in close contact, −0.11 and +0.15 respectively , one would say they have a weak favourable electrostatic interaction. For the same reason, the two hydrogens in the non-chlorine FIMVAY analogue should

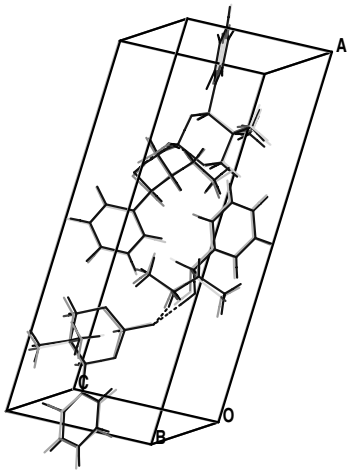


Figure 4.20: Match between VASP optimized structures of FIMVAY and its non-chlorine analogue

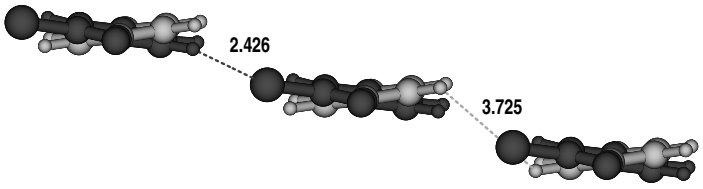


Figure 4.21: Comparison of the orientations of the phenyl rings connected to the cyclic phosphoric rings of the VASP optimized structures of FIMVAY and its non-chlorine analogue

have a weak repulsive electrostatic interaction (+0.15 is the typical value of the ESP charge of a hydrogen connected to a phenyl ring). We have searched the Cambridge Structural Database for the number of occurrences, N, of a chlorine connected to a phenyl ring having a close contact with a hydrogen connected

to another phenyl ring. The results are displayed in Figure 4.22. If the spatial distribution of nonbonded atoms were random, the probability of finding an atom at a certain distance increases with the surface area of a sphere whose radius is that distance. In practice however, one often finds a more or less well defined peak superimposed on a monotonic rising (albeit not with the square of the distance) background.[62] In our case a small peak is found at a distance of 3.1Å. This is an indication for the existence of a weak energetically favourable interaction for such a contact. In the experimental structure of FIMVAY this close contact is at 3.07Å, in agreement with the statistical analysis. However, the close contact in the VASP optimized structure is at a much smaller value (2.43Å). This seems to suggest that the description of this type of interaction using the current approximations to the density functional is not accurate enough, or perhaps that, in spite of the large unit cell, it could require a better k-sampling than the one used here (Γ k-point).

Finally, the influence of better k-space sampling on the relative lattice energies was investigated. We discussed the results at the Γ k-point in detail since it allows for a better comparison with the DMol^3 calculations, which are only possible at the Γ k-point. Table 4.19 shows the absolute and relative energies of the diastereomers of CPA-E and ClCPA-E with k-space sampling of $(2 \times 2 \times 2)$, which yields two inequivalent k-points in the irreducible part of the Brillouin zone [e]. The atomic positions are directly taken from the calculation previously described (i.e., optimized using only Γ k-point for the k-sampling and keeping the cell dimensions fixed at the experimental values).

For comparison the energies with the k-space sampling at Γ, and experimental energies are given.

The better sampling has a small effect (0.5 kcal/mol or smaller) on all structures except FIMVAY. Where FIMVAY is the most stable structure at Γ (2.3 kcal/mol more stable than FIMTUQ), at $(2 \times 2 \times 2)$ k-space sampling FIMVAY is the least stable structure (2.8 kcal/mol less stable than FIMTUQ). The absolute energy of the FIMVAY structure is raised by only 0.06 kcal/mol going from $(2 \times 2 \times 2)$ to $(3 \times 3 \times 3)$ k-space sampling, indicating that the $(2 \times 2 \times 2)$ k-space sampling is sufficient. Inspection of the forces on the atoms at the $(2 \times 2 \times 2)$

[e] $2 \times 2 \times 2$ in principle yields eight k-points, but time reversal symmetry reduces this number to four k-points and point group symmetry reduces this number to two k-points

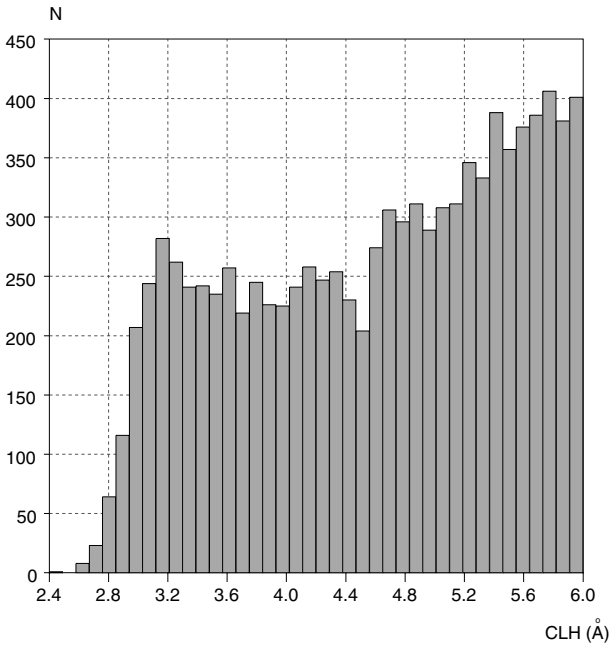


Figure 4.22: The number of occurrences, N, of a chlorine connected to a phenyl ring having a close contact with a hydrogen connected to a phenyl ring in the CSD as a function of the distance of the close contact CLH (Ångstrom units)

level on all six structures shows that they are very small except for the FIMVAY structure.

The three chlorine containing diastereomers (FIMVAY, FIMTUQ and SUMWEC) were optimized at the $(2 \times 2 \times 2)$ k-space sampling. The results are displayed in Table 4.20. Lack of computer resources prevented us from performing the optimization of all six structures at the $(2 \times 2 \times 2)$ level, but we feel this is justified in view of the small forces found at the $(2 \times 2 \times 2)$ level for all except the FIMVAY structure.

As expected, the structural and energetical changes of FIMTUQ and SUMWEC upon optimization are negligible. On optimization FIMVAY is only 0.16 kcal/mol less stable than FIMTUQ at the $(2 \times 2 \times 2)$ level. At the $(2 \times 2 \times 2)$ level, the

Table 4.19: Absolute (eV) and relative (kcal/mol) energies of the diastereomers of CPA-E and ClCPA-E with k-space sampling of two points (2 × 2 × 2), at atomic positions taken from the optimization with the k-space sampling at Γ (1 × 1 × 1) and cell dimensions fixed at the experimental values. For comparison the energies with the k-space sampling at Γ and experimental energies are given. All calculations were performed at the density functional level of theory with the VASP program.

Refcode	E_{2x2x2}	ΔE_{2x2x2}	ΔE_{Γ}	$\Delta E_{Exp.}$
FIMVEC	−352.763625	0	0	0
FILGAI	−352.783868	−0.5	−1.1	−0.1
FIMVAY analogue	−352.675767	2.0	1.4	–
FIMTUQ	−351.127602	0	0	0
SUMWEC	−351.158707	−0.7	−1.2	0
FIMVAY	−351.007007	2.8	−2.3	−2.8

SUMWEC structure is the most stable of the diastereomeric salts of ClCPA-E, being 0.7 kcal/mol more stable than FIMTUQ. Surprisingly, the relative stability of the non-chlorine containing FIMVAY analogue does not change at the (2×2×2) level. However, the structural agreement of the experimental and VASP optimized FIMVAY structure is improved going from Γ k-point to (2 × 2 × 2) k-space sampling. The dihedral angle defining the orientation of the phenyl ring containing the chlorine with respect to the cyclic phosphoric ring changes from −124.6° at the Γ k-point to −135.7° at the (2 × 2 × 2) k-space sampling. The latter is only 2° in error with the experimental value, −137.7°. The close contact of the chlorine connected to the phenyl ring with the hydrogen connected to its symmetry copy is also improved. At the (2 × 2 × 2) k-space sampling it

Table 4.20: Absolute (eV) and relative (kcal/mol) energies of the diastereomers of ClCPA-E optimized at k-space sampling of two points ($2 \times 2 \times 2$) with cell dimensions fixed at the experimental values. For comparison, the energies with the k-space sampling at Γ and experimental energies are given. All calculations were performed at the density functional level of theory with the VASP program.

Refcode	E_{2x2x2}	ΔE_{2x2x2}	ΔE_{Γ}	$\Delta E_{Exp.}$
FIMTUQ	−351.128663	0	0	0
SUMWEC	−351.159572	−0.7	−1.2	0
FIMVAY	−351.121896	0.2	−2.3	−2.8

is 2.97Å compared to 3.07Å in the experimental structure. The results for the dihedral are summarized in Table 4.21.

Table 4.21: Dihedral defining the orientation of the chlorine-containing phenyl ring with respect to the cyclic phosphoric ring for the structures FIMVAY and the non-chlorine containing FIMVAY analogue, experimental versus VASP optimized values at k-space sampling Γ *and* $2 \times 2 \times 2$.

structure	*Dihedral Angle (in Degrees)*
FIMVAY, Experiment	-137.7
FIMVAY, VASP optimized at Γ	-124.6
FIMVAY, VASP optimized at $2 \times 2 \times 2$	-135.7
FIMVAY analogue, VASP optimized at Γ	-140.9

Table 4.22: Absolute (Hartrees) and relative (kcal/mol) energies of the diastereomers of ClCPA-E, with atomic positions optimized at the density functional level of theory with the DMol^3 program, keeping cell dimensions fixed at the experimental values, compared with experimental energies.

Refcode	E_Γ	ΔE_Γ	$\Delta E_{Exp.}$
FIMTUQ	−2050.349184	0	0
SUMWEC	−2050.352856	−2.3	0
FIMVAY	−2050.351756	−1.6	−2.8

4.4.5 Lattice energy minimizations using DMol^3

Table 4.22 shows the absolute and relative energies of the diastereomers of ClCPA-E, with atomic positions optimized at the density functional level of theory with the DMol^3 program,[42] while keeping cell dimensions fixed at the experimental values, compared with experimental energies. These density functional calculations were performed within the generalized gradient approximation (GGA), employing the PW91 functional[48] (the same as used with VASP) and the double numeric with polarization (DNP) basis-set. Due to limited computer resources only the diastereomers ClCPA-E were optimized. Each optimization took more than six months of computer time on our local machine (a Silicon Graphics Power Challenge).

In agreement with experiment, FIMVAY is calculated to be more stable than FIMTUQ. However, SUMWEC is calculated to be the most stable diastereomer, though experimentally it is equally stable as FIMTUQ. Figure 4.23 shows the match between the DMol^3 optimized and experimental structures of the diastereomers of ClCPA-E with CSD reference codes: FIMTUQ, SUMWEC and FIMVAY. The agreement between calculated and experimental structures is very good with the exception of the hydrogen positions, as mentioned earlier.

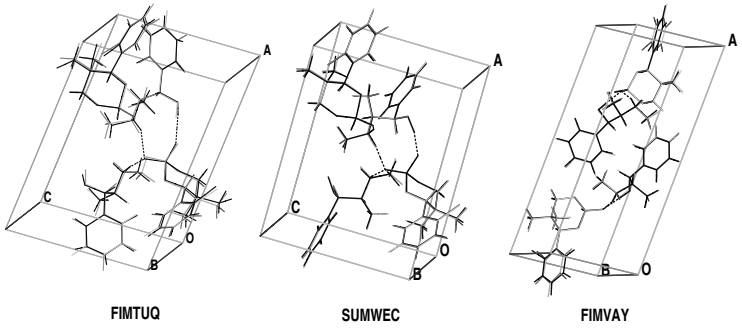


Figure 4.23: Match between DMol^3 optimized (grey) and experimental (black) structures, with CSD refcodes: FIMTUQ, SUMWEC, FIMVAY

Table 4.23 shows the comparison between the DMol^3 relative energies (kcal/mol) of the diastereomers of ClCPA-E, calculated respectively at the geometries optimized with the VASP and DMol^3 packages. The effect of the different geometries on the relative energies is negligible (largest difference 0.2 kcal/mol). This suggests that the difference in relative stability order for the diastereomers of ClCPA-E found with VASP and DMol^3 is due to differences in the way the energy is calculated. Obvious differences between the two methods are the all-electron calculation for DMol^3 versus the pseudopotential calculation of VASP and the use of different basis-sets: atom centred numerical basis functions with DMol^3 versus plane waves with VASP.

Table 4.23: Comparison between the relative energies (kcal/mol) of the diastereomers of ClCPA-E, calculated respectively at the geometries optimized with the VASP and DMol^3 packages (both using only Γ for the k-space sampling). The energies were calculated at the density functional level of theory with the DMol^3 package.

Refcode	$\Delta E_{VASP-Geom.,\Gamma}$	$\Delta E_{DMol^3-Geom.,\Gamma}$
FIMTUQ	0	0
SUMWEC	−2.3	−2.3
FIMVAY	−1.8	−1.6

Table 4.24: Relative energies (kcal/mol) of the diastereomers of ClCPA-E optimized at various levels of k-space sampling with cell dimensions fixed at the experimental values, calculated with the SIESTA, VASP and DMol^3 programs. For SIESTA a column is added containing the relative energies of the simultaneous optimization of atomic coordinates and cell dimensions at the 2x2x2 k-space level. For comparison, the experimental energies are given.

Program	*Refcode*	ΔE_{Γ}	ΔE_{2x2x2}	ΔE_{3x3x3}	ΔE_{2x2x2} $_{Opt.Cell}$	$\Delta E_{Exp.}$
SIESTA						
	FIMTUQ	0	0	0	0	0
	SUMWEC	−1.6	−1.0	−1.0	−1.1	0
	FIMVAY	−1.8	0.1	0.1	0.8	−2.8
VASP						
	FIMTUQ	0	0			0
	SUMWEC	−1.2	−0.7			0
	FIMVAY	−2.3	0.2			−2.8
DMol^3						
	FIMTUQ	0				0
	SUMWEC	−2.3				0
	FIMVAY	−1.6				−2.8

4.4.6 Lattice energy minimizations using SIESTA

Table 4.24 shows the relative energies of the diastereomers of ClCPA-E optimized at various levels of k-space sampling, with cell dimensions fixed at the experimental values, calculated with the SIESTA program.[43,44] For comparison, the VASP, DMol^3 and experimental relative energies are shown. For SIESTA a column is added containing the relative energies of the simultaneous optimization of atomic coordinates and cell dimensions at the $(2 \times 2 \times 2)$ k-space level. The SIESTA calculations were performed within the generalized gradient approximation (GGA), employing the PBE functional[49] and the double polarized basis-set.

The SIESTA results are in good agreement with the VASP results, with

differences in relative energies no bigger than 0.5 kcal/mol at the Γ k-point and 0.3 kcal/mol at the $(2 \times 2 \times 2)$ k-space level. With both VASP and SIESTA, the FIMVAY diastereomer is calculated to be the most stable at the Γ k-point. Whereas with DMol^3, the SUMWEC diastereomer is calculated to be the most stable, with a difference in relative energy between DMol^3 and SIESTA of 0.7 kcal/mol.

The SIESTA relative energies do not significantly change going from $(2\times2\times2)$ to $(3 \times 3 \times 3)$ k-space level. This is in agreement with our conclusion from the VASP results, that the $(2 \times 2 \times 2)$ k-space sampling is sufficient. As with VASP, the agreement between the SIESTA and experimental relative energies becomes worse when going from the Γ k-point to $(2 \times 2 \times 2)$ and higher k-space sampling. Figure 4.24 shows the match between the SIESTA optimized structures at $(2 \times 2 \times 2)$ k-space level and experimental structures of the diastereomers of ClCPA-E with CSD reference codes: FIMTUQ, SUMWEC and FIMVAY. The agreement between calculated and experimental structures is very good with the exception of the hydrogen positions, as was the case with VASP and DMol^3 optimized structures.

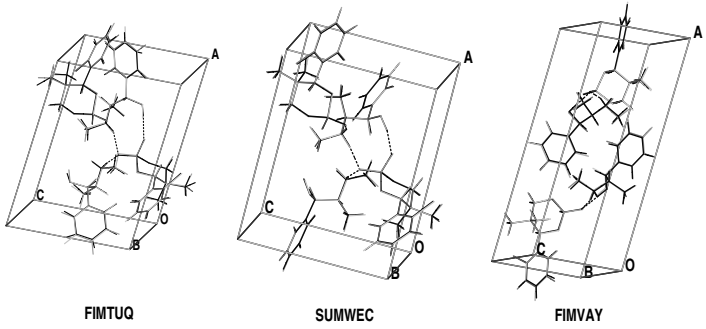


Figure 4.24: Match between SIESTA optimized structures at $(2 \times 2 \times 2)$ k-space level (grey) and experimental (black) structures, with CSD refcodes: FIMTUQ, SUMWEC, FIMVAY

When atomic coordinates and cell dimensions are simultaneously optimized at the $(2\times2\times2)$ k-space level the difference between the experimental and calculated

Table 4.25: Experimental and SIESTA optimized lattice parameters of the diastereomers of ClCPA-E. The atomic coordinates and cell dimensions were simultaneously optimized at the 2x2x2 k-space level. The last column shows the percentual increase in volume going from experimental to calculated structures.

Refcode		a (Å)	b (Å)	c (Å)	β (*Degrees*)	*Vol.* (%)
FIMVAY						
	Exp.	19.659	7.208	7.921	100.728	
	SIESTA	19.814	7.296	8.073	100.492	4.1
FIMTUQ						
	Exp.	13.842	8.381	9.812	99.672	
	SIESTA	14.099	8.453	10.015	99.708	4.7
SUMWEC						
	Exp.	13.414	8.079	10.334	91.674	
	SIESTA	13.626	8.227	10.549	92.378	5.6

results increases. The experimentally most stable ClCPA-E diastereomer, is again calculated to be the most unstable. One reason for this may be an inability of the current GGA functionals to describe ring-ring interactions properly.

Table 4.25 shows the experimental and SIESTA optimized lattice parameters of the diastereomers of ClCPA-E. The atomic coordinates and cell dimensions were simultaneously optimized at the 2x2x2 k-space level. The calculated a,b and c values are always bigger with respect to experiment (1-2 %). As a result, the calculated cell volume is increased with 4-5 % with respect to experiment.

4.4.7 The relative importance of separate interactions studied with DMol^3

To determine the relative importance of the separate interactions that play a role in the relative stabilities of the crystal packings with hydrogen-bonded chains of translational and screw symmetry, we performed a series of single point energy calculations on the modified DMol^3 optimized structures of the two best repre-

sentatives of these packings: FIMVAY and FIMTUQ. As discussed in previous sections, the relative abundances of the crystal packing with hydrogen-bonded chains of screw symmetry and translational symmetry, and the *ab initio* calculations on model systems of these chains, suggest that the hydrogen-bonded chains with screw symmetry are much more stable. In the case of FIMVAY and FIMTUQ this stability order is reversed with respect to experiment. We designed modifications in such a way that the contribution of separate interactions on the total relative stability order could be studied. These modifications are of two types, in some cases combined:

- Cell dimensions a and c increased to 100Å, thus excluding interactions in the a and c directions
- stripping parts of the system and capping cut bonds with hydrogen atoms

Table 4.26 shows the relative energies of comparable subsystems of the diastereomers FIMVAY and FIMTUQ. In cases where the subsystem contains chlorine, relative energies of the subsystems are also given, but with chlorine substituted by hydrogen ($E_{Rel}^{noCl^{(*)}}$). The entry labelled A gives the relative energies of the unmodified FIMVAY and FIMTUQ crystals with the energy minimized with respect to atomic coordinates at the density functional level of theory employing the PW91 functional[48] and the DNP basis-set with the DMol^3 package.[42] The unmodified FIMVAY crystal is 1.6 kcal/mol more stable than the FIMTUQ crystal, in agreement with the experimental value of 2.1 kcal/mol. When we increase the lattice parameters a and c to 100Å(system labelled B), the subsystems of FIMVAY and FIMTUQ are almost equally stable. The interactions in the a and c directions are all ring-ring interactions. Apparently they are more favourable for FIMVAY than for FIMTUQ.

The subsystem B of FIMVAY consists of two hydrogen-bonded chains with translational symmetry. These two chains are related by a screw-axis transformation (see Figure 4.25). In the FIMVAY subsystem C, these two chains have been separated in such a way that they are equally distant from each other as from their copies in neighbouring cells, thus the interaction between these two chains is switched off. This subsystem is 2.9 kcal/mol less stable than the FIMTUQ subsystem B. Apparently the two hydrogen-bonded chains with translational symmetry in FIMVAY have a favourable interaction of 2.7 kcal/mol ($2.9 - 0.2$). This

Table 4.26: Relative energies in kcal/mol calculated with the DMol[3] program for the diastereomers of CPA-E and ClCPA-E and a number of its modifications

Label	*Description*	E_{Rel}	$E_{Rel}^{noCl(*)}$
A	FIMVAY minus FIMTUQ	-1.6	4.8
B	FIMVAY minus FIMTUQ, a=c=100$\AA$	0.2	6.3
C	B, two FIMVAY H-bonded chains separated	2.9	9.0
D	C, chains stripped of phenyl ring	3.5	
E	D, chains stripped of cyclic phosphorus rings	5.0	
$B-C$	two FIMVAY H-bonded chains separated minus conjoined	2.7	
F	FIMVAY, structure E, separated minus joined	0.7	
G	FIMVAY, ephedrine phenyls only, separated minus joined	0.9	
H	FIMVAY chlorine substituted phenyl rings only a=b=c=100$\AA$ minus a=c=100$\AA$, b=7.208$\AA$	3.2	
I	H with chlorines substituted by hydrogens	-2.1	

$^{(*)}$ pertaining to same structures, with Cl substituted by H

is confirmed by comparing directly the energies of the FIMVAY subsystems with joined chains versus separated chains, entry $B-C$ in Table 4.26. This interaction of the two FIMVAY hydrogen-bonded chains apparently has two contributions: the phenyl-phenyl interaction of the ephedrines and the electrostatic interaction of the highly charged centres in the core of these two chains. This contention is supported by the calculation on subsystems G and F. The subsystems G are the ephedrine phenyl rings of FIMVAY only, separated or joined. The joined FIMVAY subsystem is 0.7 kcal/mol more stable than the separated one. This is in contradiction with reports in the literature on the disability of a number of DFT

functionals to describe this interaction as favourable, as discussed earlier. The subsystems F are the FIMVAY H-bonded chains with translational symmetry, separated and joined. In addition, all the phenyl rings and cyclic phosphoric rings have been stripped away. In this way, only the electrostatic interaction between the highly charged centres in the core of the hydrogen-bonded chains is studied. It appears that the two FIMVAY chains can gain 0.7 kcal/mol by electrostatic interaction.

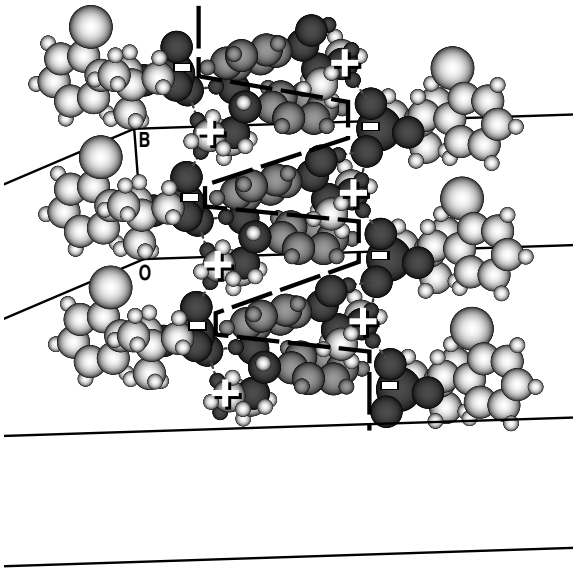


Figure 4.25: The two translationally-symmetric hydrogen-bridged chains of the FIMVAY diasteromer separated by a dashed line. The interaction between these two chains has contributions from the phenyl-phenyl interaction (green) and the electrostatic interaction of the highly charged centres in the core of the two chains (black)

If we strip the phenyl rings of the subsystems C of FIMVAY and FIMTUQ, we get subsystems D, and if we strip even the cyclic phosphoric rings we get subsystems E. These subsystems are now identical in composition to the expanded

(see figures 4.8 and 4.9) and small model systems (see figures 4.11 and 4.12) we used for our *ab initio* study of three complex hydrogen bonded chains of translational symmetry (FIMVAY) and screw symmetry (FIMTUQ). In this study we found the screw-symmetric H-bonded chains to be much more stable (by roughly 8 kcal/mol per complex unit, see Table 4.6). Subsystem D of FIMVAY is 3.5 kcal/mol less stable than subsystem D of FIMTUQ. Subsystem E of FIMVAY is 5.0 kcal/mol less stable than subsystem E of FIMTUQ. This is in qualitative agreement with results for the model calculations. We did not expect that both types of calculation would yield absolute quantitative agreement, since in the model, systems were optimized with no interactions in the a and c directions and lacked the phenyl-rings on the cyclic phophoric ring and on the ephedrine.

Comparing the relative energies of the subsystems A to C with the same systems with the chlorine substituted by hydrogen (E_{Rel} versus $E_{Rel}^{noCl^{(*)}}$), we find that the relative energies all shifted by 6.0−6.5 kcal/mol in favour of the FIMTUQ subsystems. Clearly the substitution of chlorine by hydrogen has a locally destabilising effect. This is confirmed by the relative energies of the FIMVAY subsystems consisting of only the phenyl rings containing the chorine with a=b=c=100 $\AA$ versus the same system with b=7.208$\AA$ (the experimental value), labelled H. We find that the first system, where the chloro-phenyl rings still have interactions, is 3.2 kcal/mol more stable. Looking at the relative energy of the same systems when chlorine is substituted by hydrogen (entry I), we find that the first system, where the phenyl rings still have interactions, is 2.1 kcal/mol less stable. This is in agreement with the results of the VASP calculations for the diastereomers of CPA-E. Here we found that the no-chlorine FIMVAY analogue was 3.6 kcal/mol less stable than the no-chlorine FIMTUQ analogue: FIMVEC (see Table 4.18).

4.4.8 Comparison of the absolute lattice energies

The lattice energy is defined as the amount of energy released when a mole of gaseous ions or molecules are brought together from infinite separation to form a crystal. For diastereomeric salts the gas phase energy is the same. Therefore differences in lattice energies are equivalent to differences in the solid state energy. Upto this point, when we discussed relative lattice energies, we actually discussed

Table 4.27: Comparison of the absolute lattice energies (kcal/mol) calculated with the DREIDING, DMAREL and DMol³ for the diastereomers of ClCPA-E.

Refcode	$E_{DREIDING}$	E_{DMAREL}	E_{DMol^3}
FIMTUQ	−172.2	−97.0	−129.8
SUMWEC	–	−95.6	−132.1
FIMVAY	−163.1	−100.3	−131.4

relative solid state energies, with the exception of the DMAREL calculations. In order to calculate the absolute lattice energies we need to subtract the energy of the gas phase ions from the solid state energy.

Table 4.27 shows the comparison of the absolute lattice energies (kcal/mol) calculated with DREIDING, DMAREL and DMol³ for the diastereomers of ClCPA-E. DREIDING was used with charge set 4. The DMAREL results are uncorrected for conformational energy differences of the ions, and using unscaled multipoles. The lattice energies for DREIDING, DMAREL and DMol³ show large variations, they are roughly 170, 100 and 130 kcal/mol respectively. This not necessarily means that the relative lattice energies should also show large variations. It does show, however, the difference in balance between the electrostatic and van der Waals contributions to the intermolecular interaction energy for these methods.

4.5 Conclusions

Ab initio structure optimizations on complex model chains with translational and screw symmetry yielded a good structural fit between experiment and calculation. The energy difference beteen the model complexes with translational and screw symmetry was largely conserved in the three complex model chains. The screw-symmetric model chain was found to be around 8 kcal/mol more stable per complex than its translationally symmetric counterpart. This in agreement with experiment, since experimentally, seven crystal structures of the diastereomers of CPA with ephedrine have the screw-symmetric chain, and only one has the translational symmetry.

The calculated stability order of the diastereomeric salts with the DREIDING force field and various type of charges was wrong. Moreover, the calculated relative energies of the diastereomeric salts are wrong by one order of a magnitude.

The calculated stability order of the diastereomeric salts with the DMAREL program is dependent on which multipole scaling factor is used and whether the energies are corrected for the conformational energy differences of the rigid bodies. The conformational energy differences were found to be at least of the same magnitude as the packing energy differences. Although the calculated relative energies are of the correct magnitude, the correct stability order of the diastereomeric salts was only once predicted correctly for the diastereomers of ClCPA-E, when calculations were made with unscaled multipoles and no correction for conformational energy differences of the rigid bodies. However, the errors in the calculated lattice parameters are large when using unscaled multipoles (up to 12% for the structure FIMVAY). The smallest errors in the calculated lattice parameters were found with a multipole scaling constant of 0.7.

The calculated stability order of the diastereomeric salts with the VASP/SIESTA programs at Γ point in k-space is in qualitative and quantitative agreement with experiment. The non-chlorine analogue of the experimentally most stable diastereomer, FIMVAY, was calculated to be the least stable non-chlorine diastereomer. This is probably in agreement with experiment, since this crystal packing was not observed experimentally. The stability of the diastereomer FIMVAY is probably due to an electrostatically favourable interaction between a chlorine connected to a phenyl ring and hydrogen connected to a phenyl ring. With improved

k-space sampling, however, this agreement with experiment is partly destroyed, with FIMVAY being calculated to be as stable as its diasteromer FIMTUQ.

The calculated stability order of the diastereomeric salts FIMTUQ and FIMVAY with the DMol^3 program at Γ is in qualitative and quantitative agreement with experiment and with the VASP calculations at the Γ point. The third diasteromer, SUMWEC, however is incorrectly predicted to be the most stable of the three. We do not know any reason for the disagreement between the calculations with DMol^3 and VASP/SIESTA despite the similarity of these methods.

Relative energy calculations on subsystems of the diastereomers FIMTUQ and FIMVAY confirmed the higher stability of the isolated screw-symmetric chains over those with translational symmetry. The relative energies of the packed chains in the complete crystals may be reversed, as is the case with FIMTUQ and FIMVAY, due to interactions other than those considered in the model chains. In the cases of FIMTUQ and FIMVAY there is an electrostatic interaction between a chlorine connected to a phenyl ring and hydrogen connected to a phenyl ring that stabilizes the translationally symmetric chain structure FIMVAY with respect to the screw symmetry chain structure FIMTUQ, where this interaction is absent.

Ab initio calculations predict all diastereomers of CPA-E and ClCPA-E to lie roughly within a 1 kcal/mol range. Experiment puts this range at 2 kcal/mol (for ClCPA-E). This shows that the calculations are at least 1 kcal/mol in error. New exchange-correlation functionals, that better describe the long range dispersion and ring-ring interactions, are probably required to improve the agreement between calculation and experiment. Although disagreement between DMol^3 and VASP/SIESTA also shows other sources of error.

4.6 Acknowledgements

The investigations reported in this paper were supported by the Netherlands Organization for Chemical Research (NWO-CW) within the framework of the PPM/CMS-c crystallization project. This project is a Dutch research collaboration with academic and industrial partners, focusing on pre-competitive research into modelling, packing, morphology and industrial crystallization of organic compounds. Special thanks are due to Dr. J.H. Noordik and Rob Meier for the

initialization of this project, and to DSM for sponsoring this project and in particular we would like to thank Rob Meier and Betty Coussens for their interest in this project. DSP acknowledges support by Grants No. DOE 8371494, and No. DEFG 02/96/ER 45439, and from the Basque Government (Programa de Formación de Investigadores). The SIESTA calculations presented in this paper were performed at the NCSA computational resources.

References and Notes

[1] De Camp, W.H., Chirality, 1 (1989) 2.

[2] F. Leusen, *Rationalization of racemate resolution: a molecular modelling study.* Nijmegen: Ph.D. thesis, 1993.

[3] Leusen, F.J.J., Noordik, J.H., Karfunkel, H.R., Tetrahedron, 49 (1993) 5377.

[4] Leusen, F.J.J., Bruins Slot, H.J., Noordik, J.H., van der Haest, A.D., Wynberg, H., Bruggink, A., Recl. Trav. Chim. Pays-Bays, 111 (1992) 111.

[5] L. Hansen, *Structural Investigations of Diastereomeric Salts.* The Royal Danish School of Pharmacy: Ph.D. thesis, 1997.

[6] Allen, F.H., Bellard, S.A., Brice, M.D., Cartwright, B.A., Doubleday, A., Higgs, H., Hummelink, T., Hummelink-Peters, B.G., Kennard, O., Motherwell, W.D.S., Rodgers, J.R., Watson, D.G., Acta Crystallogr., Sect. B, 35 (1979) 2331.

[7] Moers, F.G., Smits, J.M.M., Beurskens, P.T., Ariaans, G.J.A., Zwanenburg, B.Leusen, F.J.J., Bruggink, A., J. Chem. Cryst., 24 (1994) 179.

[8] Del Re, G., J. Chem. Soc. London, 1958 (1958) 4031.

[9] Gasteiger, J., Marsili, M., Tetrahedron, 36 (1980) 3219.

[10] Mulliken, R.S., J. Chem. Phys., 23 (1955) 1833.

[11] R. Bader, *Atoms in Molecules - A Quantum Theory.* Oxford: Oxford University Press, 1990.

[12] Hirshfeld, F.L., Theor. Chim. Acta, 49 (1977) 129.

[13] Breneman, C.M., Wiberg, K.B., J. Comp. Chem., 11 (1990) 361.

[14] Besler, B.H., Merz, K.M., Kollman, P.A., J. Comp. Chem., 11 (1990) 431.

[15] Bayly, C.I., Cieplak, P., Cornell, W.D., Kollman, P.A., J. Phys. Chem., 97 (1993) 10269.

[16] Schaftenaar, G., Noordik, J.H., J. Comp.-Aided Molecular Design, 14 (2000) 233.

[17] Mayo, S.L., Olafson, B.D., Goddard III, A., J. Phys. Chem., 94 (1990) 8897.

[18] Molecular Simulations Inc., 9685 Scranton Road, San Diego CA, USA, Cerius User Guide, March 1997., 1997.

[19] Stone, A.J., Chem. Phys. Lett., 83 (1981) 233.

[20] Price, S.L., Stone, A.J., Alderton, M., Molec. Phys., 52 (1984) 987.

[21] Ponder, J.W., Richards, F.M., J. Comp. Chem., 8 (1987) 1016.

[22] Mooij, W.T.M., van Duijneveldt, F.B., van Duijneveldt-van de Rijdt, J.G.C.M., van Eijck, B.P., J. Phys. Chem., A103 (1999) 9872.

[23] Mooij, W.T.M., van Eijck, B.P., Kroon, J., J. Phys. Chem., A103 (1999) 9883.

[24] Coombes, D.S., Price, S.L., Willock, D.J., Leslie, M., J. Chem. Phys., 100 (1996) 7352.

[25] Willock, D.J., Price, S.L., Leslie, M., Catlow, C.R.A., J. Comp. Chem., 16 (1995) 628.

[26] Schaftenaar, G., Noordik, J.H., J. Comp.-Aided Molecular Design, 14 (2000) 123.

[27] Hohenberg, B., Kohn, W., Phys. Rev. B., 136 (1964) 864.

[28] Kohn, W., Sham, L.J., Phys. Rev. A., 136 (1988) 1133.

[29] Pople, J.A., Binkley, J.S., Seeger, R., Int. J. Quant. Chem. Symp., 10 (1976) 1.

[30] Møller, C., Plesset, M.S., Phys. Rev., 46 (1934) 618.

[31] Pudzianowski, A.T., J. Phys. Chem., 100 (1996) 4781.

[32] Hobza, P., Šponer, J., Reshel, T., J. Comp. Chem., 16 (1995) 1315.

[33] Krystyán, S., Pulay, P.,, Chem. Phys. Lett., 229 (1994) 175.

[34] Pérez-Jordá, J.M., Becke, A.D., J. Chem. Phys., 233 (1995) 134.

[35] Meijer, E.J., Sprik, M.S., J. Chem. Phys., 105 (1996) 8684.

[36] Rovira, C., Novoa, J.J., J. Phys. Chem. B, 105 (2001) 1710.

[37] Kresse, G., Hafner, J., Phys. Rev. B, 47 (1993) 558.

[38] Kresse, G., Furthmüller, J., Comput. Mat. Sci., 6 (1996) 15.

[39] Kresse, G., Furthmüller, J., Phys. Rev. B, 55 (1996) 11169.

[40] Vanderbilt, D., Phys. Rev. B, 41 (1990) 7892.

[41] Kresse, G., Hafner, J., J. Phys.: Condens. Matter, 6 (1994) 8245.

[42] Delley, B., J. Chem. Phys., 92 (1990) 508.

[43] Sanchez-Portal, D., Ordejon, P., Artacho, E., Soler, J.M., Int. J. of Quant. Chem., 65 (1997) 453.

[44] Artacho, E., Sanchez-Portal, D., Ordejon, P., Garcia, A., Soler, J.M., Phys. Stat. Sol. (b), 215 (1999) 809.

[45] Ordejón, P., Drabold, D.A., Grumbach, P.M., Martin, R.M., Phys. Rev. B., 48 (1993) 14646.

[46] Mauri, F., Galli, G., Car, R., Phys. Rev. B., 47 (1993) 9973.

[47] Troullier, N., Martins, J.L., Phys. Rev. B., 43 (1991) 1993.

[48] Perdew, J.P., Wang, Y., Phys. Rev. B, 45 (1992) 13244.

[49] Perdew, J.P., Burke, K., Ernzerhof, M., Phys. Rev. Lett., 77 (1996) 3965.

[50] Monkhorst, H.J., Pack, J.D., Phys. Rev. B, 13 (1976) 5188.

[51] Ewald, P.P., Ann. Phys., 64 (1921) 253.

[52] Boys, S.F., Benardi, F., Mol. Phys., 19 (1970) 553.

[53] Becke, A.D., J. Chem. Phys., 98 (1993) 5648.

[54] Frish, M.J., Trucks, G.W., Head-Gordon, M., Gill, P.M.W., Wong, M.W., Foresman, J.B., Johnson, B.G., Schlegel, H.B., Robb, M.A., Replogle, E.S., Gomperts, R., Andres, J.L., Raghavachari, K., Binkley, J.S., Gonzalez, C., Martin, R.L., Fox, D.J., Defrees, D.J., Baker, J., Stewart, J.J.P., Pople, J.A., Gaussian94. Gaussian, Inc., Carnegie Office Park, Pitssburgh, PA 15106.

[55] Del Bene, J.E., Person, W.B., Szczepaniak, K., J. Chem. Phys., 99 (1995) 10705.

[56] Ley-Yeh, H., Williams, D.E., Acta Crystallogr., Sect. A, 36 (1980) 277.

[57] Cox, S.R., Williams, D.E., J. Comp. Chem., 2 (1981) 304.

[58] Furthmüller, J., Käckell, P., Bechstedt, F., Kresse, G., Phys. Rev. B., 61 (2000) 4576.

[59] Louie, S.G., Froyen, S., Cohen, M.L., Phys. Rev. B, 26 (1982) 1738.

[60] King-Smith, R.D., Payne, M.C., Lin, J.S., Phys. Rev. B, 44 (1991) 13063.

[61] Kresse, G., unpublished

[62] Rowland, R.S., Taylor, R., J. Chem. Phys., 100 (1996) 7384-7391.

Chapter 5

The gas phase chemistry of the methyl carbamate radical cation

This chapter has been reproduced with kind permission from Schaftenaar G., Postma R., Ruttink P.J.A., Burgers P.C., McGibbon G.A., Terlouw J.K., Int. J. Mass Spectrom., 100 (1990) 521.

5.1 Summary

The unimolecular chemistry of the methyl carbamate radical cation, $H_2NCOOCH_3^{+\bullet}$, **1**, has been investigated by a combination of mass spectrometry based experiments (metastable ion (MI), collisional activation (CA), collision-induced dissociative ionization (CIDI), neutralization-reionization (NR) spectrometry, ^2H, ^{13}C and ^{18}O isotopic labeling, appearance energy (AE) measurements), and ab initio molecular orbital calculations, executed at the SDCI/6-31G**//4-31G level of theory and corrected for zero-point energies.

These calculations indicate that besides ionized methyl carbamate there are at least seven other equilibrium structures including distonic ions and hydrogen-bridged radical cations. The most stable isomer is the hydrogen-bridged species $[H_2N\text{-}CH{=}O \cdots H \cdots O{=}C\text{-}H]^{+\bullet}$ which is best viewed as the carbenium ion $H_2N\text{-}CH\text{-}OH^+$ interacting with the formyl dipole. The related species $[H_2N\text{-}C{=}O \cdots H \cdots O{=}CH_2]^{+\bullet}$ in which the hydroxyaminocarbene ion $H_2N\text{-}C\text{-}OH^{+\bullet}$ interacts with the formaldehyde dipole is also a stable species. This hydrogen-bridged radical cation is the key intermediate in the spontaneous unimolecular dissociations of methyl carbamate ions.

Experimentally, the metastable molecular ions form two sets of products, namely, $H_2N\text{-}CH\text{-}OH^+ + HCO^{\bullet}$ (the components of the most stable isomer) and $[CH_2{=}O \cdots H \cdots NH_2]^{+\bullet} + CO$. The minimum energy requirement paths have

been located by ab initio calculations and the reactions follow multistep isomerizations. In the first step, $H_2NCOOCH_3^{+\bullet}$, **1**, isomerizes via a 1,4-hydrogen shift to the distonic ion $H_2NC(OH)OCH_2^{+\bullet}$, **2**, which then rearranges to the hydrogen-bridged radical ion $[H_2N\text{-}C{=}O \cdots H \cdots O{=}CH_2]^{+\bullet}$. The incipient formaldehyde molecule can then donate a hydrogen to the C atom of H_2NCOH, followed by loss of $HCO^{\bullet}$ or it can accept the hydroxyl hydrogen to form a $^{\bullet}CH_2OH$ radical; this radical then migrates within the electrostatic field of the $H_2N^{+}{=}C{=}O$ ion towards the N atom to form the complex $[H_2C{=}O \cdots H \cdots NH_2C{=}O]^{+\bullet}$. This latter species, which can be viewed as a formaldehyde and a CO molecule interacting with $NH_3^{+\bullet}$ lies in a shallow potential well only and sheds C=O to produce $[CH_2{=}O \cdots H \cdots NH_2]^{+\bullet}$, as observed experimentally.

It is stressed that only with the aid of high level ab initio calculations could the above mechanisms be elucidated.

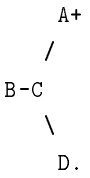
5.2 Introduction

From the inception of organic mass spectroscopy it has been realized that many a molecular ion may undergo more or less complex rearrangement reactions. This is especially true for reactions taking place in the microsecond time frame, for the metastable ions. However, rearrangement reactions may also dominate in the submicrosecond time frame and the resulting ionic products may lead to intense (and even to base) peaks in the conventional mass spectra [1]: 25 years ago it was stated [2] that "the occurrence of such rearrangements is one of the problems which complicate exceedingly the simple interpretation of ionisation and fragmentation processes and is moreover a serious obstacle to the determination of the constitution of the original compound from the mass spectrum". However, even before the above statement was made, Biemann [3] wrote, "The formation of rearrangement peaks of high intensity thus requires a certain arrangement of atoms in the molecule, which makes such peaks very useful for the interpretation of mass spectra". Biemann goes on to write "A deeper knowledge of the mechanism of these rearrangements is extremely important for this purpose". Indeed, it has been repeatedly shown that a good understanding of ionic processes, such as rearrangement reactions, can lead to a more confident interpretation of mass spectra of unknowns. In this respect it is worthy to note that in 1983 the name of

this very Journal[a] was changed to include the word Processes instead of Physics. Often metastable ions [4] are chosen for mechanistic studies because (i) their internal energy, although not well defined, lies in a narrow range, (ii) the nominal mass of the neutral expelled can be measured, (iii) when in doubt the structure of the neutral can be assessed by collision-induced dissociative ionization (CIDI), a variant of neutralization-reionization mass spectrometry (NRMS) [5], and (iv) the kinetic energy release, even when in the submillielectronvolt range, can be accurately measured. Even within a homologous series of small metastable ions, strikingly different dissociation reactions can be encountered. Let us consider the following: methyl formate ions, H-C(=O)-$OCH_3^{+\bullet}$, produce CO (for a detailed computational and experimental study see ref. [6]), methyl acetate ions, CH_3-C(=O)-$OCH_3^{+\bullet}$, form $CH_2OH^{\bullet}$ in admixture with $CH_3O^{\bullet}$ (for a recent review see ref. [7]), ionized methyl propanoate, CH_3CH_2-C(=O)-$OCH_3^{+\bullet}$, generates $CH_3O^{\bullet}$ but not $CH_2OH^{\bullet}$ [8] and methyl butyrate ions, $CH_3CH_2CH_2$-C(=O)-$OCH_3^{+\bullet}$, produce $CH_3^{\bullet}$ and C_2H_4 (by the McLafferty rearrangement; see, for example, ref. [9]). Ethyl acetate, CH_3-C(=O)-$OCH_2CH_3^{+\bullet}$, amazingly, produces water molecules [10].

From recent work, it has become clear that much can be learned about reaction mechanisms and ion structures by integration of results obtained from state of the art ab initio calculations and from mass spectrometry based experiments [7,11,12]. Previous results [11,12] have shown that ab initio calculations can lead to a proper description of the isomerization/dissociation pathways provided that a split-valence basis set is employed for geometry optimizations and that the final relative energies include the effects of polarization functions and of correlation energy. Using this approach it has become clear that many radical cations exist for which the neutral counterpart is unknown and that such ions are key intermediates in the rearrangement/dissociation pathways of many organic radical cations [7]. Of particular interest are the so-called distonic ions [13], hydrogen-bridged radical cations [14] and ion-dipole complexes [11,15]. Distonic ions are those radical cations which have the charge and radical sites at separate heavy atoms, e.g. $^{\bullet}CH_2O^+H_2$ as opposed to $CH_3OH^{+\bullet}$. In larger distonic ions (e.g. A^+-B-C$^{\bullet}$-D) often the charged moiety can migrate to the radical centre, thereby generating another distonic ion

[a] *International Journal of Mass Spectrometry and Ion Processes*



which in turn can shift its ionized part to the radical site [7,13] (B-C$^{\bullet}$-D-A^{+}). Such a sequence can rationalize otherwise mechanistically problematic reactions. In such a rationalization it is assumed that the rearrangements occur stepwise as in the Whitmore mechanism in solution carbenium ion chemistry [16]. Another class of ions which are increasingly being invoked to rationalize fragmentations are hydrogen-bridged radical cations, formally proton-bound molecule radical pairs A $\cdots$ H^{+} $\cdots$ BH$^{\bullet}$ which, unlike their even-electron counterparts A $\cdots$ H^{+} $\cdots$ B, have only been scantly studied. In such radical ions a hydrogen shift can lead to another hydrogen-bridged cation $^{\bullet}$HA $\cdots$ H^{+} $\cdots$ B which in turn can undergo a similar reaction.

Such a sequence has been invoked to explain the seemingly complex unimolecular chemistry of ionized 1,2-ethanediol [14(a)] and 1,2-propanediol [14(b)]. With respect to the above it is of interest to note that high level ab initio calculations have convincingly shown that formation of $^{\bullet}$CH$_2$OH from ionized methyl acetate proceeds via *both* distonic ions and hydrogen-bridged radical cations as shown below [12(a)].

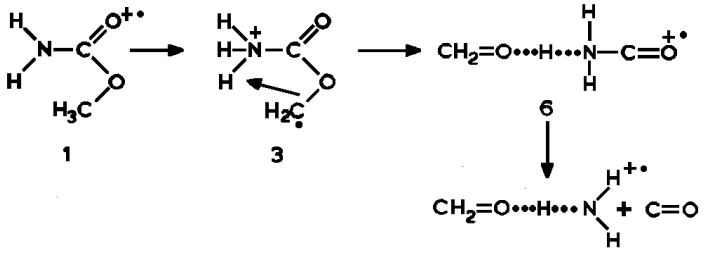
$$H_3C\text{-}C(=O)\text{-}O\text{-}CH_3^{+\bullet} \rightarrow H_3C\text{-}C^{+}(\text{-}OH)\text{-}O\text{-}CH_2^{\bullet} \text{ (distonic)}$$

$$\rightarrow CH_3C{=}O \cdots H \cdots O{=}CH_2^{+\bullet} \text{ (H-bridged)}$$

$$\rightarrow CH_3C{=}O^{+} + CH_2OH^{\bullet}$$

In a recent computational and experimental study [12(b)] on the CH_5NO potential energy surface we provided evidence that loss of CO from ionized methyl carbamate $H_2NCOOCH_3^{+\bullet}$, **1**, leads to the hydrogen-bridged radical cation $[CH_2{=}O \cdots H \cdots NH_2]^{+\bullet}$, formally the ionized formaldehyde-ammonia associated molecule and not the conventional species $CH_3ONH_2^{+\bullet}$ (by CO extrusion).

On the basis of a variety of mass spectrometric experiments we proposed that the reaction proceeds as follows:



and we note here that this sequence too includes a distonic $\rightarrow$ hydrogen-bridged transformation by intramolecular hydrogen bonding (chelation). However, we stressed that without computational support this proposal must remain speculative. Thus we embarked upon a computational study of the methyl carbamate rearrangement/dissociation reactions and some further experiments which we report here. Our ab initio calculations show that loss of CO occurs via a pathway completely different from that shown above. Also a rationale is given for the dominant reaction in the microsecond time frame, viz. loss of $HCO^{\bullet}$.

Before presenting the evidence which now leads us to reject the above mechanism it may be useful to summarize briefly the grounds upon which it was based.

(i) The daughter ions are $[CH_2{=}O \cdots H \cdots NH_2]^{+\bullet}$ and not $CH_3ONH_2^{+\bullet}$ or $^{\bullet}CH_2ONH_3^{+}$.

(ii) $H_2N\text{-}C(=O)\text{-}^{18}O\text{-}^{13}CH_3^{+\bullet}$ specifically loses $^{12}C^{16}O$

(iii) Electron capture by the (rearranged) molecular ions leads to complete breakdown indicating that the neutralized species lie(s) in a shallow well only as expected for structures akin to **3** and **6**. Major neutral products are $CH_2{=}O$ and $^{\bullet}CH_2OH$, *not* $CH_3O^{\bullet}$, consistent with structures **3** and **6**.

(iv) The CH_5NO daughter ions also fall apart upon neutralization to, among other products, $^{\bullet}CH_2OH$, *not* $CH_3O^{\bullet}$; the CH_2D_3NO daughter ions gener-

ated from the OCD_3 -labelled ester form $^{\bullet}CD_2OH$ and $^{\bullet}CD_2OD$ radicals in a ratio of 2:1.

All the above observations lend strong support for the above mechanism.

5.3 Experimental

Metastable ion (MI), collisional activation (CA) and neutralization reionization (NR) mass spectra [5] were recorded on VG Analytical ZAB-2F and ZAB-E instruments as described in ref. [12(b)]. The NR mass spectra were obtained using xenon for neutralization and O_2 for reionization; O_2 was also used as the target gas in the CA experiments. IE and AE values were obtained using monochromatic electrons as described in ref. [17]. AE values for metastable peaks were obtained by the comparative method described in ref. [19] using the KRATOS MS 902S instrument at the University of Ottawa (Professor J.L. Holmes). The labelled esters were synthesized on a microscale from the appropriately labelled alcohols (CD_3OH, CD_2HOH, $CH_3^{18}OH$ and $^{13}CH_3OH$, isotopic purity greater than 95 %, Merck Sharp and Dohme) and urea using zinc acetate as a catalyst [18]. For the synthesis of the doubly labelled ester H_2N-C(=O)-^{18}O-$^{13}CH_3$, a methanol sample (Amersham, U.K.) containing $^{13}CH_3^{18}OH$ (10 %), $^{13}CH_3^{16}OH$ (80 %) and $^{12}CH_3^{16}OH$ (10 %) was used.

5.4 Results and discussion

Loss of CO is, among metastable ions, a minor reaction path only and most of the metastable ions dissociate by loss of $HC{=}O^{\bullet}$ (95 %). Loss of $HC{=}O^{\bullet}$, was not investigated in our previous study and we deal with it here. Firstly, what is the structure of the daughter ions? A 1,5-hydrogen shift in **6** followed by loss of $HC{=}O^{\bullet}$ would lead to H_3N^+-CH=O, **a**. However, formation of the isomeric carbenium ion H_2N-C^+H-OH, **b**, may be energetically more attractive since ion **b** has two powerful electron-donating groups attached to the carbenium centre. The proton affinity of formamide H-C(=O)-NH_2 has been measured [20(a)], yielding an enthalpy of formation of 123 kcal mol^{-1} for the protonated molecule but its structure, **a** or **b**, has not been determined. However, our theoretical calculations

show that ion **b** is considerably more stable than ion **a** (by 14.6 kcal mol^{-1}) and thus the protonation of formamide probably yields CH_4NO^+ ion **b**.

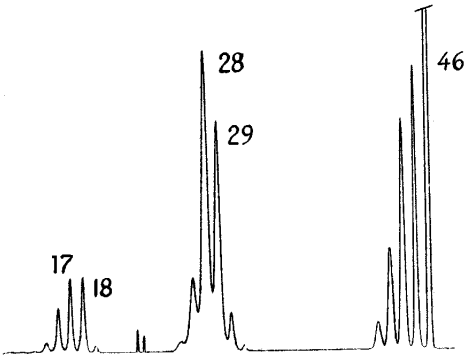


Fig. 5.1: CA mass spectrum of the CH_4NO^+ ions generated by loss of $HCO^\bullet$ from ionized methyl carbamate, $H_2NCOOCH_3^{+\bullet}$.

The CA mass spectrum of the CH_4NO^+ ions generated by loss of $HC{=}O^\bullet$ is shown in Fig. 5.1. Spontaneous reactions lead to m/z 18, NH_4^+, which (see Table 5.1) is by far the least energy demanding reaction. There are intense peaks at m/z 29, 28 and also at m/z 17. The CA mass spectrum is compatible with ion **a**, which is expected to dissociate to $H\text{-}C{=}O^+$ (which can fragment further to $CO^{+\bullet}$) and to $NH_3^{+\bullet}$. However, there are also clear signals at m/z 27, probably $HNC^{+\bullet}$, at m/z 16, NH_2^+, and at m/z 30 (loss of $NH_2^\bullet$) indicative of structure **b**. The NR spectrum of the CH_4NO^+ ions is shown in Fig. 5.2A. Two important observations are that (1) there is a recovery signal at m/z 46 of considerable intensity, and (2) the NR spectrum is very similar to the CA mass spectrum (m/z 18 is much less intense in the NR spectrum as this peak arises from the slow dissociation of the reionized species). These observations indicate that the neutralized species is stable for at least 0.5 μs and this argues against ions **a** which are not expected to survive neutralization.

Further evidence for this assignment comes from the NR mass spectrum of the $CH_2D_2NO^+$ ions generated by $D\text{-}C{=}O^\bullet$ loss (no $H\text{-}C{=}O^\bullet$ loss is observed) from the OCD_3 labelled ester, shown in Fig. 5.2B. If we were dealing with type **a** ions

Table 5.1: Enthalpies (from ref. 20(a)) of products for CH_4NO^+ dissociations

m/z	Products	$\Sigma\Delta H_f$ (kcal mol^{-1})
17	$NH_3^{+\bullet}$ + $HCO^\bullet$	234
18	NH_4^+ + CO	125
28	$HCNH^+$ + H_2O	168
29	HCO^+ + NH_3	186
29	$HCNH_2^+$ + $OH^\bullet$	267
30	$HCOH^{+\bullet}$ + $NH_2^\bullet$	275

these ions could be H_2DN^+-CD=O and/or HD_2N^+-CH=O. However, since m/z 19 is very weak (cf. the intensity of m/z 17 in Fig. 5.2A) the latter structure can be discarded and so the species would have to be H_2DN^+-CD=O. This ion, after NR, should yield $CD{=}O^+$ followed by loss of $D^\bullet$ to produce peaks at m/z 30 and 28, but no peak at m/z 29 is expected, in sharp contrast with the experimental findings (Fig. 5.2B). The m/z 29 peak must then be the result of HDO loss and so about 40 % of m/z 28 in the unlabelled NR and CA spectra is not $CO^{+\bullet}$ but rather $HNCH^+$ formed by loss of H_2O. These findings are compatible with the proposal that the daughter ions are indeed H_2N-C^+H-OH, **b**, which incidentally can produce m/z 29 $HCNH_2^+$ ions via a simple cleavage reaction (see Table 5.1). That loss of H_2O is observed at all is not so surprising considering the products' enthalpies. The labelled $CH_2D_2NO^+$ daughter ions are then H_2N-CD-OD (as also evidenced by the clean shift of m/z 30 to m/z 32, $CDOD^{+\bullet}$).

It should be noted that the CA mass spectrum of H_2N-C^+H-OH (Fig. 5.1) contains weak signals at m/z 19 and m/z 31. These signals could well arise from a trace of ethanol [14(a)]: the corresponding signals are absent in the CA mass spectrum of the labelled species (not shown). We mention this because a trace amount of ethanol may render the interpretation of the AE m/z 46 value

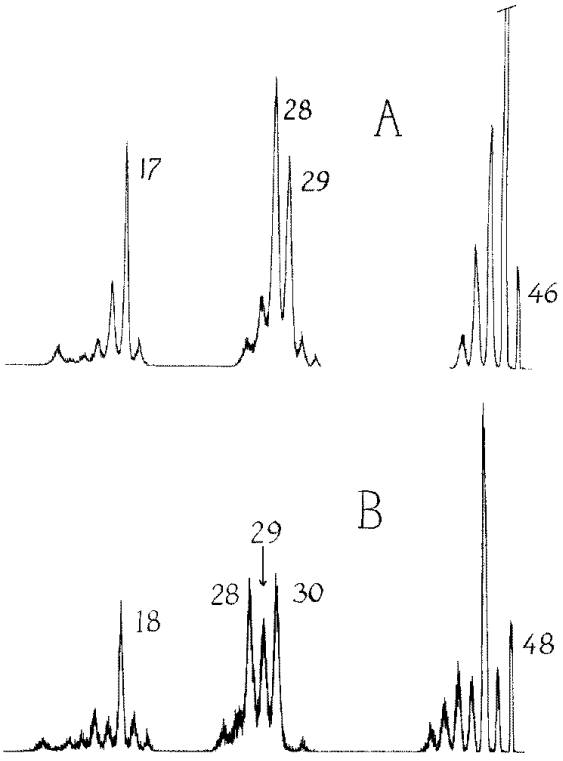


Fig. 5.2: A, [Xe, O_2] NR mass spectrum of the CH_4NO^+ ions generated by loss of $HCO^\bullet$ from ionized methyl carbamate, $H_2NCOOCH_3^{+\bullet}$; B, NR mass spectrum of the $CH_2D_2NO^+$ ions generated by loss of $DCO^\bullet$ from the labelled ester $H_2NCOOCD_3^{+\bullet}$

questionable (see below).

To obtain more information on the mechanism of $HCO^\bullet$ and CO loss the MI and CA spectra of several labelled molecules were investigated and the results are summarized as follows:

H_2N-C(=O)-O-$CD_3^{+\bullet}$	loses	$DCO^{\bullet}$	and	CO
H_2N-C(=O)-O-$^{13}CH_3^{+\bullet}$	loses	$H^{13}CO^{\bullet}$	and	^{12}CO
H_2N-C(=O)-^{18}O-$CH_3^{+\bullet}$	loses	$HC^{18}O^{\bullet}$	and	$C^{16}O$
H_2N-C(=O)-^{18}O-$^{13}CH_3^{+\bullet}$	loses	$H^{13}C^{18}O^{\bullet}$	and	$^{12}C^{16}O$

Both rearrangement/dissociation processes (loss of $HCO^{\bullet}$ and CO) are thus atom specific and the following reactions are established:

$$H_2N\text{-}C(=O)\text{-}{}^{*}O={}^{*}CH_3^{+\bullet} \rightarrow {}^{*}CH_2={}^{*}O \cdots HNH_2^{+\bullet} + C{=}O \quad (1)$$

$$\rightarrow H_2N\text{-}C^{+}H\text{-}OH + H\text{-}{}^{*}C={}^{*}O^{\bullet} \quad (2)$$

5.4.1 *Energetic measurements*

The IE and AE values were measured using monochromatic electrons (see experimental section): IE($H_2NCOOCH_3$) = 10.44 $\pm$ 0.05 eV, yielding $\Delta H_f(H_2NCOOCH_3)^+$ = 141 kcal mol^{-1} using $\Delta H_f(H_2NCOOCH_3)$ = -100 kcal mol^{-1}. AE m/z 46 was measured as 10.82 $\pm$ 0.05 eV, which is only 0.35 eV above the IE of ethanol [20(a)]. Since we had noted above that a trace of ethanol could well be present, it is conceivable that the measured value is too low. To remove this ambiguity the AE of the metastable peak for loss of $HCO^{\bullet}$ was measured by the comparative method described in ref. [19], AE = 11.20 $\pm$ 0.1 eV. The AE of the metastable peak for loss of CO was also measured by this method, AE = 11.15 $\pm$ 0.1 eV. The latter AE value was found to be identical with that obtained using monochromatic electrons (AE = 11.16 $\pm$ 0.05 eV) which gives credence to the metastable measurements. Therefore we propose that the activation energies for loss of $HCO^{\bullet}$ and CO are the same within experimental error.

The ΔH_f for H_2N-C^{+}H-OH as derived from the above AE measurement (149 kcal mol^{-1}) is much higher than that of protonated formamide (123 kcal mol^{-1}). This leads to the conclusion that the loss of $HCO^{\bullet}$ from **1** is associated with a significant energy barrier (about 1.1 eV) for the reverse reaction. It should be noted that this is not reflected in the magnitude of the kinetic energy release: $\langle T \rangle$ is only 77 meV. This point will be addressed later. The formation of the hydrogen-bridged species CH_2=O $\cdots$ H $\cdots$ $NH_2^{+\bullet}$, too, may involve a barrier for

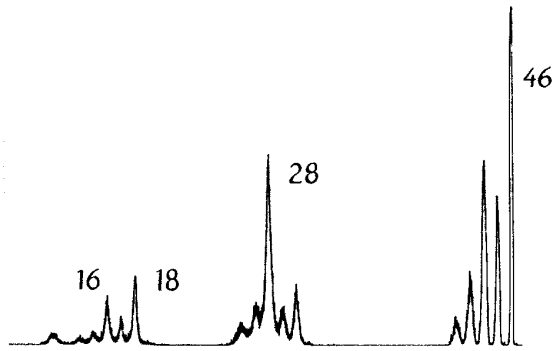


Fig. 5.3: [Xe, O_2] NR mass spectrum of the CH_2DNO^+ ions generated by loss of CD_2O from the labelled ester $H_2NCOOCD_3^{+\bullet}$.

the reverse reaction and so the experimentally determined ΔH_f (184 kcal mol^{-1}) [12(b)] must be regarded as an upper limit.

At slightly higher energies (AE = 11.24 eV) another process comes into play, namely dissociation into m/z 44 $H_2N\text{-}C^+{=}O$ ions and $\Delta m = 31$ neutrals, $^\bullet CH_2OH$ (ΔH_f = -6.2 kcal mol^{-1} [20(a)]) or $CH_3O^\bullet$ (ΔH_f = +3.7 kcal mol^{-1} [20(a)]). Depending on the structure of the neutral lost, the apparent ΔH_f of the m/z 44 $H_2N\text{-}C^+{=}O$ ions is 155 kcal mol^{-1} (for loss of $CH_3O^\bullet$)and 165 kcal mol^{-1} (for loss of $^\bullet CH_2OH$). Since ΔH_f for $H_2N\text{-}C^+{=}O$ has been established as 161 kcal mol^{-1} [20(b)] it follows that at threshold the radical eliminated is $^\bullet CH_2OH$ and not $CH_3O^\bullet$.

At still higher energies a fourth reaction, loss of $CH_2{=}O$, comes into play, for which the AE is 11.44 eV. This process could lead to the ionized carbene $H_2N\text{-}C\text{-}OH^{+\bullet}$ or to the distonic ion $H_3N^+\text{-}C^\bullet{=}O$. The NR mass spectrum of the labelled daughter ion $CH_2DNO^{+\bullet}$ formed from the OCD_3 labelled ester is shown in Fig. 5.3. It is observed that the recovery signal is the base peak and that intense signals are present at m/z 16, NH_2^+ and m/z 18, OD^+. Since the hypervalent species $H_2DN{=}C{=}O$ is not expected to survive it is proposed that the daughter ion is $H_2N\text{-}C\text{-}OD^{+\bullet}$, which upon neutralization yields the stable carbene $H_2N\text{-}C^{\bullet\bullet}\text{-}OD$. The daughter ion is thus $H_2N\text{-}C\text{-}OH^{+\bullet}$ and not $H_3N^+\text{-}C^\bullet{=}O$, in agreement with our theoretical calculations which predict that the N-C bond strength in

H_3N^+-C$^\bullet$=O is only 0.8 kcal mol^{-1} (see below). ΔH_f for H_2N-C-OH$^{+\bullet}$, barring a reverse term, is assessed at 190 kcal mol^{-1}.

The higher energy losses of CH_2=O and $^\bullet CH_2OH$ (or $CH_3O^\bullet$) give rise to prominent peaks in the CA mass spectrum [12(b)]. These reactions too appeared atom specific: the H_2N-C(=O)-^{18}O-$^{13}CH_3$ labelled ester loses exclusively $^{13}CH_2^{18}O$ and $^{13}CH_3^{18}O^\bullet$ and/or $^{13}CH_2^{18}OH^\bullet$. Thus for the collision-induced dissociations the following reactions are established:

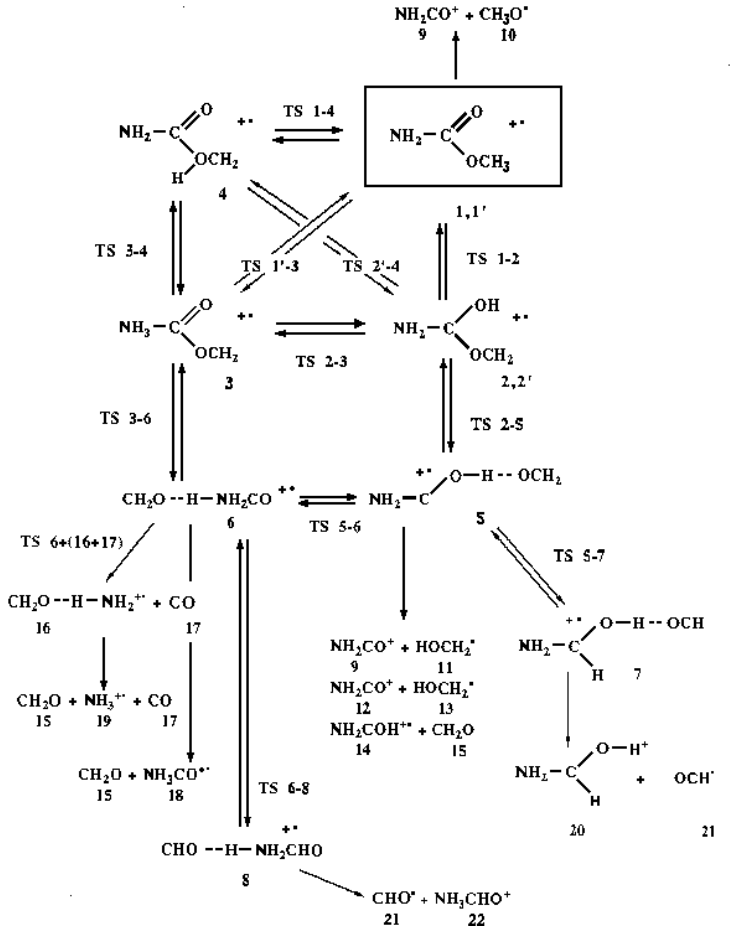
$$\begin{aligned} H_2N\text{-}C(=O)\text{-}^{18}O\text{-}^{13}CH_3^{+\bullet} &\rightarrow H_2N\text{-}C\text{-}OH^{+\bullet} + {}^{13}CH_2{=}^{18}O \\ &\rightarrow H_2N\text{-}C^+{=}O + {}^{13}CH_3^{18}O^\bullet/{}^{13}CH_2^{18}OH^\bullet \end{aligned}$$

It is remarkable that whereas the AE value for loss of $^\bullet CH_2OH$ is only slightly higher than the AE value for loss of HCO$^\bullet$ and CO, no loss of $^\bullet CH_2OH$ is observed in the MI spectrum: the signal at m/z 44 is at least a factor of 400 less intense than that at m/z 46 (loss of HCO$^\bullet$) and in addition it may well result from collision-induced dissociation by residual background gas. That the losses of CO and HCO$^\bullet$ compete in the microsecond time frame indicates that their activation energies are almost, if not exactly, the same; the latter will be so if these reactions proceed via the same rate-determining barrier (see below).

5.4.2 *Theoretical methods*

Standard LCAO-MO-SCF calculations were performed with the program GAMESS [21] employing restricted Hartree-Fock (RHF) procedures. The isomers, transition states and dissociation products which have been examined are given in Scheme 1. The geometries of the isomers were determined using analytical gradient and numerical second-derivative optimization procedures with the 4-31G basis set. These are displayed in Fig. 5.4. The calculated energies are shown in Table 5.2. The relative energies presented in Table 5.2 were obtained by performing single plus double configuration interaction (SD-CI) calculations in the 4-31G optimized geometries using the 6-31G** basis set with the help of the DIRECT CI program [22]. The SD-CI results were size consistency corrected using the formula of Pople et al. [23].

The Pople size consistency corrected results were corrected for the scaled (0.9) contribution of zero-point vibrational energies (ZPVE). The ZPVEs [24] in Table 5.2 were calculated for the 4-31G optimized geometries employing the 4-31G basis set.



Scheme 1. $C_2H_5NO_2$ isomers and transition states studied theoretically.

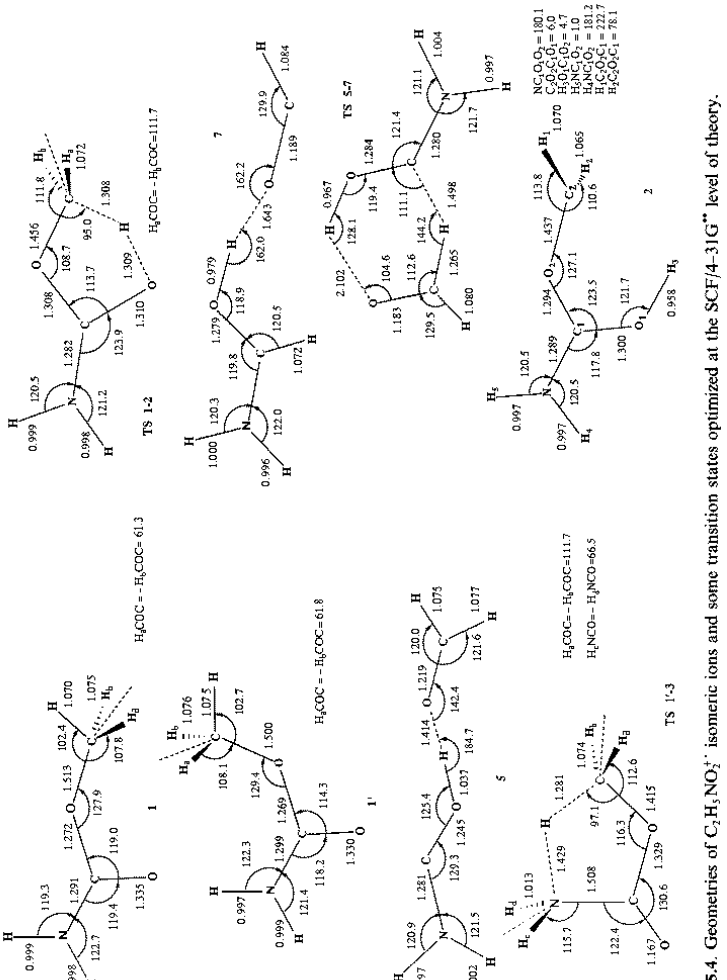


Fig. 5.4. Geometries of $C_2H_5NO_2^{+\cdot}$ isomeric ions and some transition states optimized at the SCF/4-31G** level of theory.

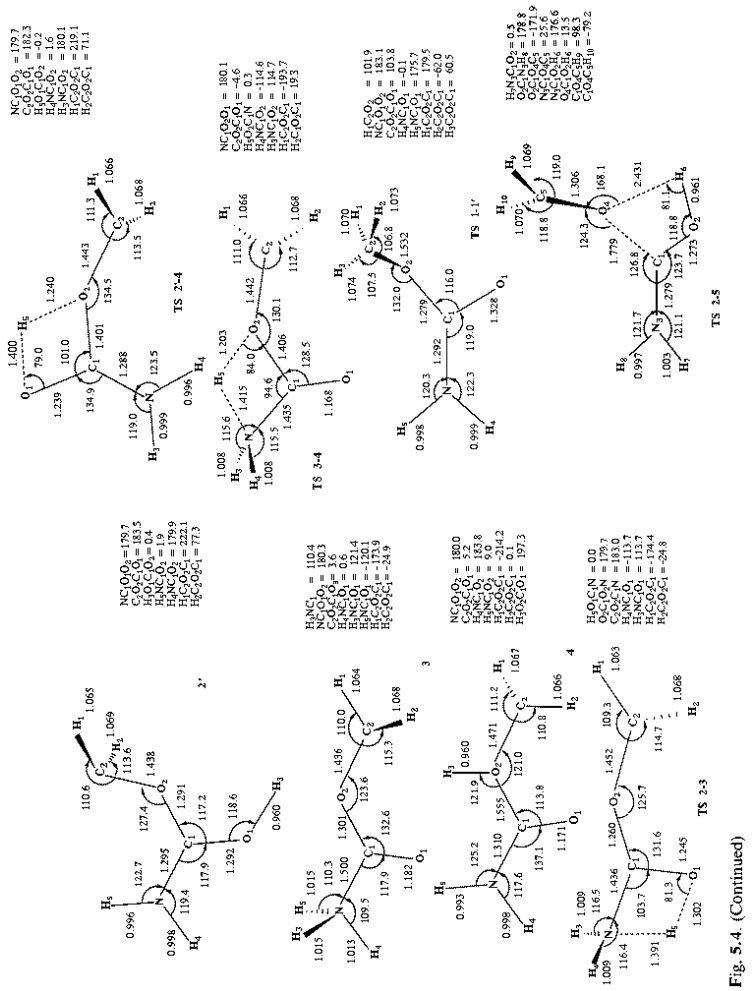


Fig. 5.4. (Continued)

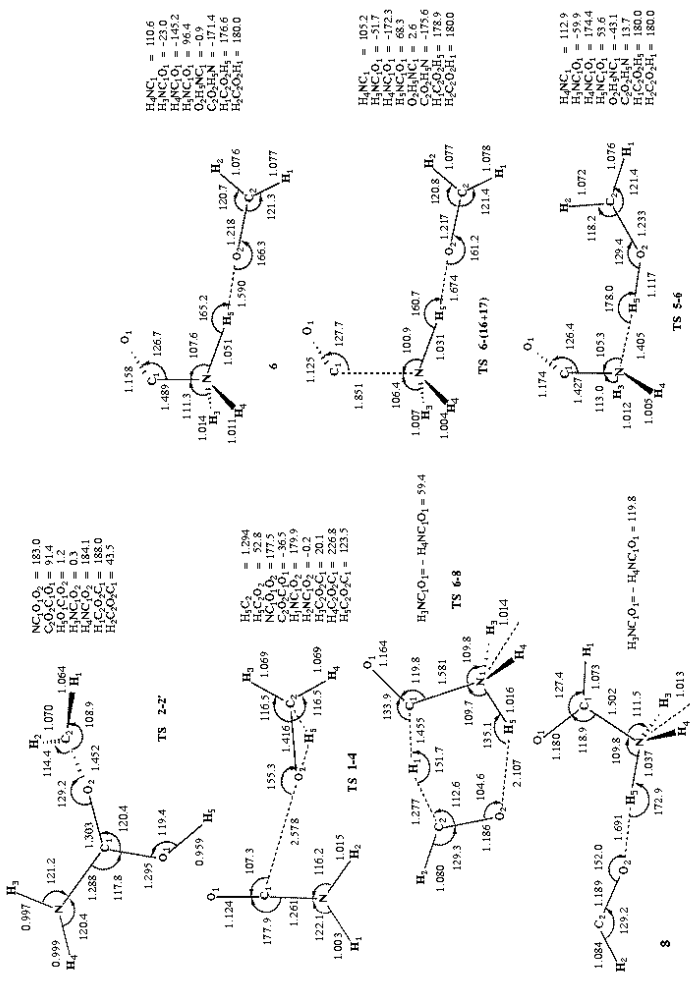
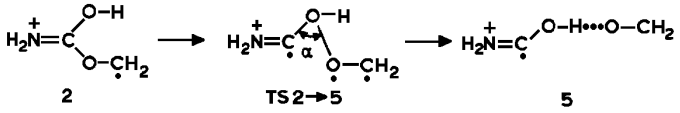


Fig. 5.4. (Continued)

5.4.3 *The unimolecular chemistry of methylcarbamate ions*

Since we believe [12(b)] that we have indications for the participation of ion **3** in the loss of CO from **1** we first calculated its relative energy and the barrier separating it from **1**. At our highest level of theory (see Table 5.2) **3** is favoured over **1** by 4.9 kcal mol^{-1}, and so the ion is thermodynamically accessible. However, the barrier for its formation (from **1′**) turned out to be surprisingly high (26.9 kcal mol^{-1}), i.e. 11 kcal mol^{-1} higher than the experimentally derived activation energy (16 kcal mol^{-1}). Considering the level of our calculations and the generally good agreement between experiment and theory at comparable levels of theory found for other systems [7,11,12] we conclude that the isomerization **1**→**3** does not in fact take place, at least not prior to the spontaneous loss of CO or that of $HCO^{\bullet}$.

How then does the loss of CO and also that of $HCO^{\bullet}$ occur? Previous theoretical calculations on methyl esters indicate that a 1,4-hydrogen shift to the oxygen atom requires only 11-15 kcal mol^{-1}. In the case of methyl carbamate TS **1**→**2** lies only 11.6 kcal mol^{-1} above **1** and isomer **2**, the distonic ion H_2N-C(OH)-$OCH_2^{+\bullet}$, is 12.6 kcal mol^{-1} more stable than **1**. Thus it is clear that the transformation **1**→**2** can take place below the minimum energy required for the losses of CO and $HCO^{\bullet}$; by contrast the transformation **1**→**3** cannot take place below the minimum dissociation energy. Hence we conclude that the distonic ion **2** is accessible. That TS **1′** →**3** is much higher in energy than TS **1**→**2** may be traced to the fact that the NCOO π-electron system in **1** is not affected by the transition **1**→**2**, whereas in TS **1′** →**3** the π-electron delocalization does not extend to the N atom. However, for ion **2** to serve a key intermediate in the losses of CO and $HCO^{\bullet}$ it must be able to undergo rearrangement reactions whose energy requirements are below that for the direct bond cleavage reaction **2** → $H_2NCOH^{+\bullet}$ + CH_2=O (**14** + **15**). An attractive pathway which satisfies this requirement involves isomerization into the very stable hydrogen-bridged isomer **5** as the first step:



A similar isomerization has been proposed to occur in the distonic ions H-C(OH)-$OCH_2^{+\bullet}$ (generated from ionized methyl formate) and CH_3-C(OH)-$OCH_2^{+\bullet}$ (generated from ionized methyl acetate): calculations [6,11,12(a)] have shown that in these ions, stretching of the formaldehydecarbene bond does not immediately lead to dissociation but to a situation where, well below the dissociation limit, isomerization can occur to the stable hydrogen-bridged isomer R-C=O $\cdots$ H $\cdots$ O=$CH_2^{+\bullet}$ (R = H or CH_3).

Using the SCF method, however, we were not able to find a saddle point for the transition state TS (**2**→**5**). We then performed a number of partial geometry optimizations for fixed values of the angle α ($\angle$OOC) as the reaction coordinate. These calculations indicated that for $\alpha = 45\,^\circ$ two SCF solutions exist: one which is characterized by a *single* C-O bond in the formaldehyde moiety (as in ion **2**) and one in which this bond is a *double* C=O bond (as in ion **5**). The two solutions correspond to valleys in the potential energy surface which are not separated by a saddle point but by an intersecting line, as indicated in Fig. 5.5.

Table 5.2: Calculated total energies (hartree), zero-point vibrational energies (kcal mol^{-1}) and relative energies (kcal mol^{-1}) for isomers and components of ionized methyl carbamate

Species	Point group	State	RHF/4-31G //4-31G	RHF/6-31G** //4-31G	Pople/6-31G** //4-31G	ZPVE 4-31G*	E_{rel}
1	C_s	$^2A'$	-282.11727	-282.54499	-283.32871	48.0	0
1′	C_s	$^2A'$	-282.11127	-282.54028	-283.32309	48.1	3.6
2	C_1	^2A	-282.12129	-282.55630	-283.34766	47.3	-12.6
2′	C_1	^2A	-282.12419			47.4	
3	C_1	^2A	-282.09446	-282.53260	-283.33567	47.5	-4.9
4	C_1	^2A	-282.07479	-282.49841	-283.29903	46.1	16.7
5	C_s	$^2A'$	-282.10309	-282.53057	-283.32719	45.3	-1.7
6	C_1	^2A	-282.06938	-282.50513	-283.31557	45.8	6.0
7	C_s	$^2A'$	-282.12184	-282.55292	-283.35072	46.1	-15.7
8	C_s	$^2A'$	-282.08043	-282.51570	-283.32376	45.3	0.4
TS 1-1′	C_1	^2A	-282.10438			47.6	
TS 2-2′	C_1	^2A	-282.11255			47.2	
TS 1-2	C_s	$^2A'$	-282.04093	-282.48412	-283.30594	45.3	11.6
TS 1′-3	C_s	$^2A'$	-282.00160	-282.45001	-283.28028	44.6	27.0
TS 2-3	C_1	^2A	-282.00813	-282.45582	-283.26723	43.9	34.5
TS 1-4	C_1	^2A	-281.98368	-282.42355	-283.24373	41.9	47.2
TS 2′-4	C_1	^2A	-282.02086	-282.45257	-283.26192	43.8	37.7
TS 3-4	C_1	^2A	-282.01800	-282.44667	-283.26090	43.9	38.5
TS 5-6	C_1	^2A	-282.05807	-282.49663	-283.30602	43.9	10.2
TS 6-(6 + 17)	C_1	^2A	-282.05463	-282.48989	-283.30957	44.0	8.0
TS 5-7	C_s	$^2A'$	-282.02916	-282.46888	-283.29304	43.5	17.9
TS 6-8	C_s	$^2A'$	-281.99844	-282.44246	-283.27881	42.8	26.1
9+10			-282.05347	-282.48274	-283.27413	43.7	30.0
9+11			-282.03356	-282.47677	-283.27966	43.3	26.1
12+13			-282.03379	-282.47395	-283.27459	44.9	30.9
14+15			-282.05404	-282.49146	-283.29055	44.5	20.5
16+17			-282.08952	-282.52336	-283.31274	40.7	2.7
15+18			-282.02327	-282.46915	-283.28732	44.6	22.6
15+19+17			-282.04349	-282.48615	-283.27749	39.3	23.4
20+21			-282.08580	-282.52626	-283.33184	44.6	-5.4
22+21			-282.04675	-282.49118	-283.30787	44.1	9.2

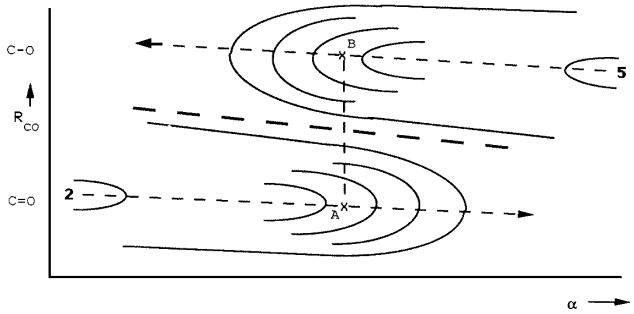


Fig. 5.5 Schematic representation of the SCF potential energy surface of the isomerization **2**→**5**

Additional SCF calculations were performed for several C-O bond lengths between the two partial SCF minima (A and B in Fig. 5.5) for a fixed angle, α = 45 °. These calculations show a discontinuity in the energy gradient of the formaldehyde C-O bond, resulting in an intersecting line between the two valleys (the bold broken line in Fig. 5.5). The discontinuity will disappear when the two SCF solutions are superimposed into a CASSCF wavefunction. In the case of a CASSCF calculation one selects an active orbital space which consists of occupied and unoccupied (virtual) orbitals. Within the active space all excitations are allowed, so a full configuration interaction (CI) calculation is done for the active orbital space chosen. In our case we need three active orbitals for the three electrons which are directly involved in the **2**→**5** isomerization reaction, viz. one electron on each of the two C atoms and one on the O atom involved in the isomerization. Using an interpolated geometry with α = 45 ° and R_{CO} = 1.45 $\AA$ as the starting point we were thus able to locate the transition state TS **2**→**5**. The calculation of the activation energy was performed at the CASSCF/4-31G level of theory and the same level of theory was used for optimizing the geometries of the ions **2** and **5**. According to these calculations the transition state TS **2**→**5** corresponds to both the breaking of the N(C)-O(C) bond in ion **2** and the breaking of one $(H_2)C{=}O$ bond in ion **5**. In TS **2**→**5** there are essentially three radical electrons. In the SCF method two of these electrons are forced to occupy the same MO. The character of the SCF wavefunction is determined by the $(H_2)C{=}O$

bond length; for intermediate CO bond lengths there is an instability with respect to localization of the radical electron on either of the two C atoms, leading to a discontinuity in the energy gradient of the CO bond.

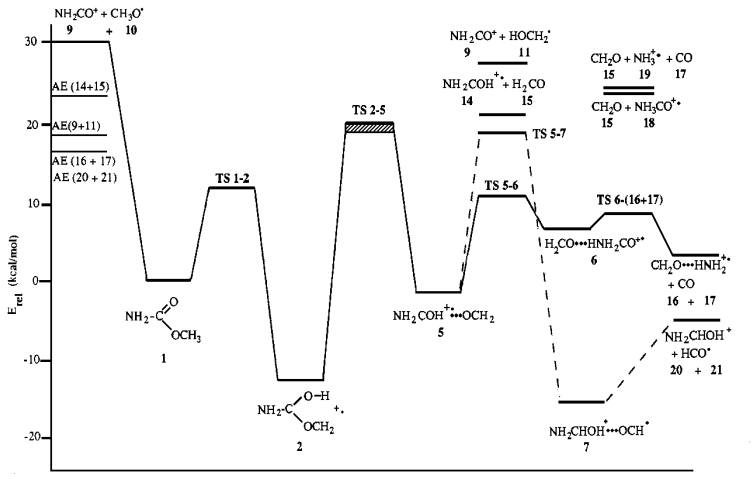


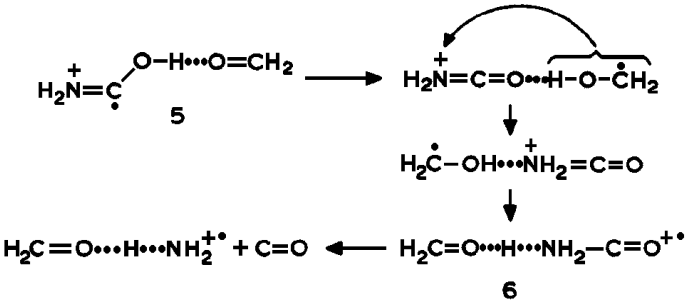
Fig. 5.6. Theoretically derived energy diagram for the dissociation reactions of ionized methyl carbamate, $H_2NCOOCH_3^{+\bullet}$.

The final results of our CASSCF calculations which, because of the large computational expense involved, had to be restricted to the CASSCF/4-31G level of theory, are as follows: ion **2**, E= - 282.14122; ion **5**, E = - 282.11974; and TS **2**→**5**, E = - 282.08514 hartree. This is equivalent to TS **2**→**5** being 35 kcal mol^{-1} higher in energy than **2**, or 22.4 kcal mol^{-1} above **1** (see Fig. 5.6). From the results in Table 5.2 it can be seen that enlarging the basis set with polarization functions generally has little effect on the activation energies. However, electron correlation effects are much more important for the transition states than for the stable isomeric ions, and their inclusion results in a decrease in the activation energies for the isomerizations. Since inclusion of the ZPVE's is expected to have the same effect we propose that the calculated activation energy for the isomerization step **2**→**5** may be an upper limit. In any case, the calculated relative energy for TS **2**→**5** is in satisfactory agreement with the experimental

observations: it lies below the dissociation limit for **2**→ $H_2NCOH^{+\bullet}$ + CH_2=O, (**14** + **15**) and it may well represent the (rate determining) second isomerization step in the mechanism for the losses of CO and $HCO^{\bullet}$.

The fate of ions **5** is intriguing: in marked contrast with the related species CH_3-O=O ··· H ··· O=$CH_2^{+\bullet}$ and H-C=O ··· H ··· O=$CH_2^{+\bullet}$ which dissociate by simple bond cleavage to their components, our calculations indicate that ion **5** can undergo two rearrangement reactions below or close to the dissociation limit. The energy diagram presented in Fig. 5.6 shows the results of our ab initio calculations. For comparison this diagram also contains the experimentally derived activation energies for loss of $HCO^{\bullet}$/CO (16 kcal mol^{-1}), CH_2=O (23 kcal mol^{-1}) and $CH_3O^{\bullet}$/$CH_2OH^{\bullet}$ (18.5 kcal mol^{-1}).

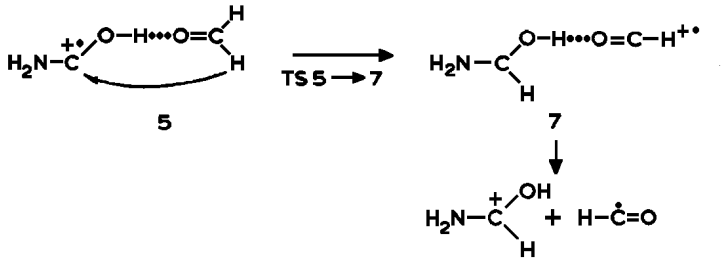
The first possibility is that the formaldehyde molecule in **5** attracts the hydroxyl hydrogen. The incipient $^{\bullet}CH_2OH$ radical subsequently moves within the electrostatic field of the H_2N=C=O ion towards the N atom to form, via TS **5**→**6**, the hydrogen-bridged species H_2C=O ··· H ··· $NH_2CO^{+\bullet}$, **6**.



Thus, in agreement with our previous proposal [12(b)], it is indeed ion **6** from which the loss of CO occurs. However, this ion is not formed via the simple pathway proposed previously, but via a more complex route. According to the calculations these ions lie in a shallow well only and they dissociate by loss of CO to produce the hydrogen-bridged product ion H_2C=O ··· H ··· $NH_2^{+\bullet}$, in agreement with experiment. It should be noted that the dissociation limit for the reaction **6** → H_3N^+-$C^{\bullet}$=O + CH_2=O is much higher in energy, i.e. the N-C

bond in **6** is much weaker than the O ··· H ··· N hydrogen-bridged ion-dipole interaction.

Alternatively, instead of accepting a hydrogen atom, the formaldehyde molecule in **5** may donate a hydrogen atom to the C atom of the NH_2-C-$OH^{+\bullet}$ ion, forming via TS **5**→**7** the most stable $C_2H_5NO_2$ isomer, H_2NCHOH^+ ··· O=C-H$^\bullet$, in which the NH_2-C-$OH^{+\bullet}$ ion interacts with the formyl dipole. Loss of HCO$^\bullet$ occurs directly therefrom:



Are the results of the calculations, as summarized in the energy diagram shown in Fig. 5.6, in agreement with experiment ? Firstly, theory predicts that metastable ions **1** will lose HCO$^\bullet$ and CO, as is observed experimentally. Secondly, theory predicts the correct structures for the daughter ions, i.e. NH_2-CH-$OH^{+\bullet}$ and $H_2C{=}O \cdots H \cdots NH_2^{+\bullet}$ respectively. Furthermore, for the labelled species, theory predicts atom-specific behaviour and it also predicts the correct isotopic positions in the products (see above), notably the observation that H_2N-C(O)-^{18}O-$^{13}CH_3$ cleanly eliminates $H^{13}C^{18}O$ and $^{12}C^{16}O$.

In addition, we had shown earlier [12(b)] that the dissociative neutralization of the OCD_3 labelled ester generated, among other products, largely $^\bullet CD_2OD$ (together with some $^\bullet CD_2OH$). Dissociative neutralization, not collisional activation, was chosen to prevent post-collisional isomerization [12(b)]. Had low energy methyl carbamate ions rearranged to **3** then mostly $^\bullet CD_2OH$ would have been generated. This is strong support for the rearrangement **1**→**2**→**5** after which dissociative neutralization to $^\bullet CD_2OD$ takes effect.

The small amount of $^\bullet CD_2OH$ formed after the neutralization indicates that in **2** the formaldehyde molecule may also migrate to the N atom, generating

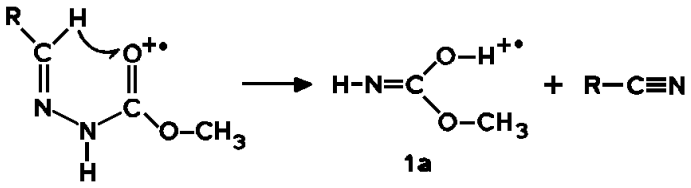
$[CH_2{=}O \cdots H \cdots NH\text{-}C\text{-}OH]^{+\bullet}$, but this was not further investigated. It should be noted that formation of $^{\bullet}CD_2OH$ cannot be ascribed to the transformation **1**→**3** as low energy ions are sampled in the NR experiment. The partially labelled ion $H_2NC(O)OCD_2H^{+\bullet}$ in the microsecond time frame loses $DCO^{\bullet}$ and $HCO^{\bullet}$ in a ratio of 2:1. This is precisely what our calculations predict, provided that no isotope effects operate in the steps **1**→**2** and **5**→**7** and so we conclude that not these steps, but the transformation **2**→**5** is rate determining as was the case with ionized methyl acetate [12(a)].

The calculated relative energies and the experimental values, where they overlap, compare favourably except for the energy difference Δ between **1** and its dissociation products $H_2N\text{-}C{=}O + {}^{\bullet}CH_2OH$, **9** + **11**: experiment (this work), Δ = 18 kcal mol^{-1} vs. theory, Δ = 26 kcal mol^{-1}. The reason for this large discrepancy is not well understood although it should be noted that, in general, dissociation energies are more difficult to establish accurately than the relative energies of a set of isomeric ions.

There is particularly good agreement between the experimentally and theoretically derived magnitude of the reverse term for loss of $HCO^{\bullet}$: 26 kcal mol^{-1} (experiment) vs. 27 kcal mol^{-1} (theory). However, the average kinetic energy release $\langle T \rangle$ is small and non-specific; $\langle T \rangle$ = 1.77 kcal mol^{-1} and so only about 7 % of the energy is released as kinetic energy. Similar effects have been observed previously in the dissociation of other hydrogen-bridged radical cations [12(b),14(a)]. Similarly the loss of CO has a calculated reverse term, E_{rev}, of about 15 kcal mol^{-1}, but $\langle T \rangle$ is only 1.07 kcal mol^{-1} (about 7 % of E_{rev}). The observation that the $CD_2{=}O \cdots D \cdots NH_2^{+\bullet}$ daughter ions formed from the OCD_3 labelled ester generate upon collisional activation a 2:1 mixture of $^{+}CD_2OH$ and $^{+}CD_2OD$ ions and upon dissociative electron capture a 2:1 ratio of $^{\bullet}CD_2OH$ and $^{\bullet}CD_2OD$ radicals can be rationalized by allowing the formaldehyde dipole to (freely) rotate around the ammonia ion, i.e. $CD_2{=}O \cdots D \cdots NH_2^{+\bullet} \rightleftharpoons CD_2{=}O \cdots H \cdots NDH^{+\bullet}$.

Finally, there are two questions related to the ionchemistry of ionized methyl carbamate which we are currently investigating. The first question concerns the possible participation of the isomeric ion $H\text{-}N{=}C(OH)\text{-}OCH_3^{+\bullet}$, **1a**, the "enol" analogue of methyl acetate, in the isomerization behaviour of $H_2N\text{-}C(=O)OCH_3^{+\bullet}$, **1**. This enol-type ion, which according to preliminary calculations is quite a sta-

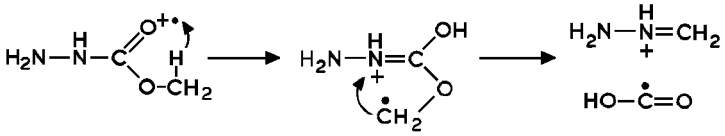
ble species, could be formed from ion **2** by a 1,4-hydrogen shift (a 1,3-hydrogen shift in ion **1** is probably energetically prohibited [12(d)]). However, if this isomerization reaction occurs then ion **1a** cannot isomerize back to **2** because the labelled ion H_2N-C(=O)-$OCD_3^{+\bullet}$ specifically loses $DCO^{\bullet}$. Ion **1a** was expected to be generated by a McLafferty-type rearrangement in RCH=N-N(H)$COOCH_3$ (R = H, CH_3), but surprisingly



1a

the MI, CA and NR spectra of the abundant m/z 75 $C_2H_5NO_2$ ions generated from these precursor molecules were very close to those of the keto ion **1**. We are currently investigating other means of generating **1a** and also establishing the height of the barriers for its (inter)conversion with ions **1** and **2** by ab initio calculations.

The second question concerns the following: if one of the amino hydrogen atoms of methyl carbamate is substituted by an NH_2 group (yielding methyl carbazate, H_2N-N(H)-$COOCH_3$) an unexpected and totally different dissociation behaviour ensues [25]. Methyl carbazate ions dissociate, among other products, to the immonium ion H_2N-N^+(H)=CH_2 by expulsion of $^{\bullet}COOH$ in a one-step process [25]. Based on a variety of experiments the following mechanism was proposed.



According to this proposal in the intermediate distonic ion, which can be considered as the analogue of **2**, a methylene unit is transferred to the charged centre. In the case of ion **2**, a similar reaction would lead to the very stable immonium ion $^{+}CH_2NH_2$ and $^{\bullet}COOH$ whose combined enthalpies ($\Sigma\Delta H_f$ [$CH_2NH_2^+$ + $^{\bullet}COOH$] = 125 kcal mol^{-1} [20(a)], are lower than those for the observed reaction products. From the start of the present study we were, in fact, surprised that methyl carbamate ions do not expel $^{\bullet}COOH$. This too, is under further investigation.

5.5 Conclusions

The results of our study provide a detailed insight into the unimolecular chemistry of ionized methyl carbamate. Although its chemistry (loss of $HCO^{\bullet}$ and CO) would, superficially, appear to be totally different from that of ionized methyl acetate (loss of $CH_3O^{\bullet}$ and $^{\bullet}CH_2OH$), the first two of the three isomerization steps are in fact the same (molecular ion $\rightarrow$ distonic ion $\rightarrow$ hydrogen-bridged ion).

However, in contrast with ionized methyl acetate, the resulting hydrogen-bridged radical cation H_2N-C-OH $\cdots$ O=$CH_2^{+\bullet}$, **5**, can undergo two further rearrangement reactions. The first process involves rearrangement to the hydrogen-bridged ion-dipole complex H_2C=O $\cdots$ H $\cdots$ NH_2-C=$O^{+\bullet}$, **6**, which leads to the loss of CO by a simple bond cleavage. The second process involves isomerization into another hydrogen-bridged ion-dipole complex, the most stable $C_2H_5NO_2^{+\bullet}$ isomer, H_2NCHOH^+ $\cdots$ O=C-H, **7**, which is the immediate precursor for loss of $HCO^{\bullet}$. Thus it appears that the participation of hydrogen-bridged ion-dipole complexes in the isomerization is even more important in the unimolecular chemistry of ionized methyl carbamate than in that of ionized methyl acetate. This is largely due to the combination of stability and flexibility in these ions.

Finally we note that, as was the case with methyl acetate [12(a)], a coherent description of the chemistry of solitary methyl carbamate ions could not have been achieved by the analysis of experimental data only.

5.6 Acknowledgements

The authors are grateful to Dr. F.P. Lossing for appearance energy measurements and to Professor J.L. Holmes for valuable discussions and for access to the KRATOS MS 902S instrument. J.K.T. thanks the National Sciences and Engineering Research Council of Canada (NSERC) and the University of Utrecht for financial support.

References and Notes

1 (a) H. Budzikiewicz, C. Djerassi and D.H. Williams, Mass Spectrometry of Organic Compounds, Holden-Day, London, 1967.
(b) F.W. McLafferty, Mass Spectrometry of Organic Ions, Academic Press, New York, 1963.

2 R.I. Reed, Quart. Rev., 20 (1966) 527.

3 K. Biemann, Mass Spectrometry: Organic Chemical Applications, McGraw-Hill, New York. 1967, p. 102.

4 (a) R.G. Cooks, J.H. Beynon, R.M. Caprioli and G.R. Lester, Metastable Ions, Elsevier, Amsterdam, 1973.
(b) K. Levsen, Fundamental Aspects of Organic Mass Spectrometry, Verlag Chemie, Weinheim, 1978.

5 (a) C. Wesdemiotis and F.W. McLafferty, Chem. Rev., 87 (1987) 485.
(b) J.K. Terlouw and H. Schwarz, Angew. Chem., Int. Ed. Engl., 26 (1987) 805; Angew. Chem., 99 (1987) 799.
(c) J.K. Terlouw, Adv. Mass Spectrom., 11 (1989) 984.
(d) J.L. Holmes, Mass Spectrom. Rev. 8 (1989) 513.
(e) F.W. McLafferty, Science, 247 (1990) 925.

6 P.C. Burgers, T. Drewello, N. Heinrich, W. Kulik, C.J. Morrow, J. Schmidt, J.K. Terlouw and H. Schwarz, in preparation.

7 P.C. Burgers and J.K. Terlouw, Spec. Period. Rep., Vol. 10, Chem. Soc., London, 1989, Chap. 2.

8 P.C. Burgers, C.E.C.A. Hop, J.K. Terlouw and J.L. Holmes, Org. Mass Spectrom., 21 (1986) 549.

9 C.E. Hudson, T. Lin and D.J. McAdoo, Org. Mass Spectrom., 22 (1987) 311.

10 (a) A.N.H. Yeo, J. Chem. Soc. D, 1970, 1154.
(b) J.L. Holmes, P.C. Burgers and J.K. Terlouw, Can. J. Chem., 12 (1981) 1805.
(c) L. Freiser-Monteiro, M.L. Freiser-Monteiro, J.L. Butler and T. Baer, J. Phys. Chem., 86 (1982) 752.

11 N. Heinrich and H. Schwarz, in J.P. Maier (Ed.), Ion and Cluster Ion Spectroscopy and Structure, Elsevier, Amsterdam, 1989, p. 329.

12 (a) N. Heinrich, J. Schmidt, H. Schwarz and Y. Apeloig, J. Am. Chem. Soc., 109 (1987) 1317 and references cited therein.
(b) P.C. Burgers, C. Lifshitz, P.J.A. Ruttink, G. Schaftenaar and J.K. Terlouw, Org. Mass Spectrom., 24 (1989) 579.

(c) W. Hehre, L. Radom, P.v.R. Schleyer and J.A. Pople, Ab Initio Molecular Orbital Theory, John Wiley, New York, 1986.
(d) B. Yates and L. Radom, J. Am. Chem. Soc., 109 (1987) 2910.

13 (a) S. Hammerum, Mass Spectrom. Rev., 7 (1988) 123.
(b) B.F. Yates, W.J. Bouma and L. Radom, Tetrahedron, 22 (1986) 6225.

14 (a) P.C. Burgers, J.L. Holmes, C.E.C.A. Hop, R. Postma, P.J.A. Ruttink and J.K. Terlouw, J. Am. Chem. Soc., 109 (1987) 7315.
(b) B.L.M. van Baar, P.C. Burgers, J.L. Holmes and J.K. Terlouw, Org. Mass Spectrom., 23 (1988) 355.
(c) P.C. Burgers, J.L. Holmes. J.K. Terlouw, B.L.M. van Baar, Org. Mass Spectrom., 20 (1985) 202.

15 (a) D.J. McAdoo, Mass Spectrom. Rev., 7 (1988) 363.
(b) R. Postma, P.J.A. Ruttink, B.L.M. van Baar, J.K. Terlouw, J.L. Holmes and P.C. Burgers, Chem. Phys. Lett., 123 (1986) 409.

16 D. Rethel and V. Gold, Carbonium Ions-An Introduction, Academic Press, New York, 1967, p. 205.

17 F.P. Lossing and J.C. Traeger, Int. J. Mass Spectrom. Ion Phys., 19 (1976) 9.

18 A.M. Paquin, Z. Naturforsch., 1 (1946) 518.

19 P.C. Burgers and J.L. Holmes, Org. Mass Spectrom., 17 (1982) 123.

20 (a) S.G. Lias, J.E. Bartmess, J.F. Liebman, J.L. Holmes, R.D. Levin and W.G. Mallard, J. Phys. Chem. Ref. Data, 17 (1988) Suppl. 1.
(b) C.E.C.A. Hop, J.L. Holmes, P.J.A. Ruttink, G. Schaftenaar and J.K. Terlouw, Chem. Phys. Lett., 156 (1989) 251 (note that this paper shows that $\Delta H_f H_2NC{=}O^+$ as quoted in ref. 20a (167 kcal mol^{-1}) is probably in error).

21 M. Dupuis, D. Spangler and J. Wendolowski. NRCC Software Catalog 1, Program No QG01, GAMESS, 1980; M.F. Guest and J. Kendrick, GAMESS User Manual, An Introductory Guide, CCP/86/1, Daresbury Laboratory, 1986.

22 V.A. Saunders and J.H. van Lenthe, Mol. Phys., 48 (1983) 923.

23 J.A. Pople, R. Seeger and N.R. Krishan, Int. J. Quantum Chem., 11 (1977) 149.

24 J.A. Pople, H.B. Schlegel, R. Krishan, D.J. DeFrees, J.S. Binkley, M.J. Frisch, R.A. Whiteside, R.F. Hout, Jr. and W.J. Hehre, Int. J. Quantum Chem., 15 (1981) 269.

25 P.C. Burgers, T. Drewello, H. Schwarz and J.K. Terlouw, Int. J. Mass Spectrom. Ion Processes, 95 (1989) 157.

Summary

Computational Chemistry Methods

Applications to racemate resolution and radical cation chemistry

This thesis deals with the application of computational models to solve real life chemical problems. Two distinct problems are tackled. First the prediction of lattice energy differences between a pair of diastereomeric salts and secondly the elucidation of the unimolecular chemistry of the methyl carbamate radical cation.

The introductory chapter deals with the background of racemate resolution via the formation of diastereomeric salts. The solubility difference between two salts of a diastereomeric pair is related to the lattice enthalpy difference in a pair. A computational model that accurately reproduces experimental lattice enthalpy differences can be used as a predictive tool for the resolution of optical isomers. In addition, we summarize the mass spectrometry techniques used for the experimental elucidation of the unimolecular chemistry of the methyl carbamate radical cation. The computational methods applied to both these problems are briefly reviewed.

Chapter 2 covers the Molden package, our pre- and post- processing program of molecular and electronic structure. This program has become central to this research. It is used to calculate a number of different types of charges in a molecule, the electrostatic potential, distributed multipoles, Z-matrix constructs to impose translational and screw type of symmetry, for interfacing with the program packages DMAREL and VASP that perform lattice energy calculations, and for manipulation of crystal structures. Other features include the calculation

/ visualization of molecular orbitals, the electron density, molecular minus atomic density and the Laplacian of the electron density.

Chapter 3 discusses the development of two new molecular point-charge models: a) charges fit to reproduce the Distributed Multipole derived electrostatic potential and b) charges fit to reproduce the quantum mechanical electrostatic potential sampled on a number of surfaces with constant electron density. In addition, we investigated the effect of the number of expansion sites in a Distributed Multipole Analysis on the quality of charges fitted to the DMA derived electrostatic potential. The inclusion of bond centers into the calculations improves the agreement between the Quantum Mechanical electrostatic potential and the DMA derived potential. The number of expansion sites needed for an accurate approximation of the QM electrostatic potential increases with increasing quality of the basis set used. Sampling on constant electron density surfaces gives a better fit between the quantum mechanical potential and the potential derived from the fitted charges, than sampling on a Van der Waals surface composed of intersecting spheres.

Chapter 4 describes Quantum Mechanical and Force Field calculations on Diastereomeric Salts. We have calculated the relative lattice energies of the diastereomers of cyclic phosphoric acid and its chlorine derivative with ephedrine, with various computational models, and compared them with experimental data. All computational models gave good structural agreement with the experiment, but only some models reproduced the experimental stability order. Calculations with the DREIDING force field in combination with several charge sets (including the charge set developed in chapter 3) failed to reproduce the experimental stability order of the diastereomers. By using distributed multipoles to model the electrostatic interactions in force field calculations, the correct stability order was reproduced in one case, but the results are very sensitive to the factor used for scaling the electrostatic interactions and to conformational energy corrections of the rigid molecular ions. Quantum mechanical calculations with the DMol^3 density functional package predict a correct stability order for one pair of diastereomers, but fail to position a third polymorph correctly. Similar calculations with the VASP and SIESTA density functional packages predict the stability order

of all diastereomers correctly at the Γ point in k-space. With better k-space sampling, however, the agreement between theory and experiment becomes less good. An experimentally unknown chlorine-free analogue of the experimentally most stable chlorine-containing diastereomer, was calculated to be the least stable. Model systems of the two types of hydrogen-bonded chains observed in a series of diastereomeric salts with screw and translational symmetry were optimized at the Hartree-Fock and density functional level. The experimentally most frequently occurring hydrogen-bonded chain with screw axis symmetry was calculated to be the most stable.

In the final chapter, the unimolecular chemistry of the methyl carbamate radical cation, $H_2NCOOCH_3^{+\bullet}$, has been investigated by a combination of mass spectrometry based experiments and ab initio molecular orbital calculations. These calculations indicate that besides ionized methyl carbamate there are at least seven other equilibrium structures including distonic ions and hydrogen-bridged radical cations. The most stable isomer is the hydrogen-bridged species $[H_2N\text{-}CH{=}O \cdots H \cdots O{=}C\text{-}H]^{+\bullet}$ which is best viewed as the carbenium ion $H_2N\text{-}CH\text{-}OH^+$ interacting with the formyl dipole. The related species $[H_2N\text{-}C{=}O \cdots H \cdots O{=}CH_2]^{+\bullet}$ in which the hydroxyaminocarbene ion $H_2N\text{-}C\text{-}OH^{+\bullet}$ interacts with the formaldehyde dipole is also a stable species. This hydrogen-bridged radical cation is the key intermediate in the spontaneous unimolecular dissociations of methyl carbamate ions. Experimentally, the metastable molecular ions form two sets of products, namely, $H_2N\text{-}CH\text{-}OH^+ + HCO^\bullet$ (the components of the most stable isomer) and $[CH_2{=}O \cdots H \cdots NH_2]^{+\bullet} + CO$. The minimum energy paths have been located by ab initio calculations. The reactions follow multistep isomerizations. It is stressed that the above mechanisms could only be elucidated with the aid of high level ab initio calculations.

In conclusion, the present day computational models for the calculation of lattice energies are still not sufficiently accurate to reliably predict and design racemate resolutions. The structures of the diastereomeric salts were correctly predicted, however. The unimolecular chemistry of the methyl carbamate radical cation could be succesfully elucidated by ab initio calculations.

Samenvatting

Computational Chemistry Methoden

Toepassingen op racemaat splitsingen en radicaal kation chemie

Dit proefschrift behandelt de toepassing van rekenmodellen voor het oplossen van chemische problemen. Twee verschillende problemen worden behandeld: de voorspelling van verschillen in roosterenergie tussen een diastereomeer zoutpaar en de opheldering van de unimoleculaire chemie van het methylcarbamaat radicaal kation.

Het eerste inleidende hoofdstuk behandelt de achtergrond van racemaatsplitsing via de vorming van diastereomere zouten. Het oplosbaarheidsverschil tussen de twee zouten van een diastereomeer paar is gerelateerd aan het verschil in roosterenergie tussen de zouten. Een rekenmodel dat nauwkeurig verschillen in experimentele roosterenergie kan reproduceren kan gebruikt worden als gereedschap bij de voorspelling van scheiding van optische isomeren. Tevens wordt een overzicht gegeven van massaspectrometrische technieken gebruikt bij de opheldering van de unimoleculaire chemie van het methylcarbamaat radicaal kation. De in beide problemen toegepaste rekenmodellen worden kort behandeld.

Hoofdstuk 2 belicht het Molden pakket, ons voor- en na-bewerkingsprogramma voor moleculaire en electronische structuur. Dit programma speelt een centrale rol in dit onderzoek. Het wordt gebruikt voor de berekening van verschillende typen ladingen in een molecuul, de electrostatische potentiaal, gedistribueerde multipolen, Z-matrix constructies voor het opleggen van translatie- en schroefassymmetrie, voor interfacing met de programma pakketten voor roosterenergieberekeningen DMAREL en VASP, en voor het manipuleren van kristalstructuren.

Andere mogelijkheden zijn de berekening/visualisatie van moleculaire orbitalen, de electronendichtheid, de moleculaire minus atomaire electronendichtheid en de Laplaciaan van de electronendichtheid.

Hoofdstuk 3 bespreekt de ontwikkeling van twee nieuwe moleculaire puntladingsmodellen: a) ladingen gefit ter reproductie van de potentiaal afgeleid van gedistribueerde multipolen en b) ladingen gefit ter reproductie van de kwantummechanische electrostatische potentiaal bemonsterd op een aantal oppervlakken van constante electronendichtheid. Tevens werd het effect van het aantal expansiecentra in een gedistribueerde multipool analyse (DMA) op de kwaliteit van de ladingen gefit aan de DMA afgeleide potentiaal onderzocht. Het meenemen van bindingscentra bij de berekeningen verbetert de overeenstemming tussen de kwantummechanische electrostatische potentiaal en de DMA afgeleide potentiaal. Het aantal expansiecentra nodig voor een nauwkeurige benadering van de QM electrostatische potentiaal neemt toe met de kwaliteit van de gebruikte basisset. Het monsteren op oppervlakken van constante electronendichtheid geeft een betere overeenkomst tussen de kwantummechanische potentiaal en de potentiaal afgeleid van de gefitte ladingen, dan het monsteren op een Van der Waals oppervlak bestaande uit overlappende bollen.

Hoofdstuk 4 behandelt kwantummechanische en krachtveld berekeningen aan diastereomere zouten. Met verschillende rekenmodellen werden de relatieve roosterenergieën van de diastereomeren van een cyclisch fosforzuur en zijn chloor derivaat met efedrine berekend en de resultaten worden vergeleken met de experimentele gegevens. Alle rekenmodellen gaven goede overeenkomst tussen berekende en experimentele structuur, echter slechts enkele rekenmodellen reproduceerden de experimenteel gevonden stabiliteitsvolgorde. Berekeningen met het DREIDING krachtveld in combinatie met verschillende ladingssets (inclusief de ladingsset ontwikkeld in hoofdstuk 3) reproduceerden niet de experimenteel gevonden stabiliteitsvolgorde van de diastereomeren. Bij gebruik van gedistribueerde multipolen voor het modelleren van de electrostatische interacties in de krachtveld berekening, wordt de correcte stabiliteitsvolgorde slechts in één geval gereproduceerd, maar de resultaten zijn erg gevoelig voor de factor gebruikt bij het schalen van de electrostatische interacties en voor de conformationele energiecor-

recties van de moleculaire ionen. kwantummechanische berekeningen met het DMol^3 density functional pakket voorspellen de correcte stabiliteitsvolgorde voor één paar diastereomeren, maar een derde polymorf wordt niet goed gepositioneerd. Vergelijkbare berekeningen met de VASP en SIESTA density functional pakketten voorspellen de correcte stabiliteitsvolgorde van alle diastereomeren bij het Γ punt in de k-ruimte. Echter, met betere bemonstering van de k-ruimte wordt de overeenkomst tussen de theorie en het experiment slechter. Een experimenteel onbekende chloorvrije analoog van de experimenteel meest stabiele chloorbevattende diastereomeer, was in de berekening het minst stabiel. Modelsystemen van de twee typen van waterstofgebrugde ketens die gevonden worden in een serie van diastereomere zouten met translatie en schroefas symmetrie werden geoptimaliseerd op het Hartree-Fock en density functional niveau. De experimenteel meest gevonden waterstofgebrugde keten met schroefas symmetrie was in de berekening het meest stabiel.

Het laatste hoofdstuk bespreekt de unimoleculaire chemie van het methyl carbamaat radicaal kation, $H_2NCOOCH_3^{+\bullet}$, onderzocht met een combinatie van massaspectrometrie experimenten en ab initio moleculaire orbitaal berekeningen. Deze berekeningen geven aan dat behalve geïoniseerd methylcarbamaat er tenminste zeven andere evenwichtsstructuren zijn, waaronder distonische ionen en waterstofgebrugde radicaal kationen. Het meest stabiele isomeer is het waterstofgebrugde $[H_2N\text{-}CH{=}O \cdots H \cdots O{=}C\text{-}H]^{+\bullet}$ dat het best gezien kan worden als het carbenium ion $H_2N\text{-}CH\text{-}OH^+$ in interactie met de formyl dipool. Het gerelateerde isomeer $[H_2N\text{-}C{=}O \cdots H \cdots O{=}CH_2]^{+\bullet}$ waarin het hydroxyaminocarbeen ion $H_2N\text{-}C\text{-}OH^{+\bullet}$ interactie heeft met de formaldehyde dipool is ook een stabiel ion. Dit waterstofgebrugde radicaal kation is een belangrijk intermediair voor de spontane unimoleculaire dissociatie-reacties van methylcarbamaat ionen. Experimenteel vormen de metastabiele moleculaire ionen twee sets van produkten, namelijk, $H_2N\text{-}CH\text{-}OH^+ + HCO^\bullet$ (de componenten van het meest stabiele isomeer) en $[CH_2{=}O \cdots H \cdots NH_2]^{+\bullet} + CO$. De minimumenergiepaden zijn gevonden met behulp van ab initio berekeningen. Benadrukt wordt dat de opheldering van de bovengenoemde reactiemechanismes slechts mogelijk was met behulp van hogekwaliteits ab initio berekeningen.

Concluderend kan gesteld worden dat de huidige rekenmodellen voor de berekening van roosterenergieën nog steeds niet accuraat genoeg zijn om betrouwbaar voorspellingen van racemaatsplitsingen te kunnen doen. De structuren van de diastereomere zouten werden echter wel goed voorspeld. De unimoleculaire chemie van het methylcarbamaat radicaal kation kon succesvol opgelost worden met behulp van hogekwaliteits ab initio berekeningen.

Dankwoord

Tot slot zou ik iedereén willen bedanken die op de één of andere manier aan de totstandkoming van dit proefschrift heeft bijgedragen. Op de eerste plaats ben ik natuurlijk dank verschuldigd aan mijn promotores Ad van der Avoird en Elias Vlieg. Zij hebben het aangedurfd een project onder hun hoede te nemen dat al drie jaar in volle gang was. Daarnaast wil ik Jan Noordik bedanken, onder wiens leiding ik aan dit werk begonnen ben. Met name Jan's oprechte belangstelling voor mij als persoon en mijn werk en zijn geloof in mijn kunnen zijn mij zeer tot steun geweest. Robert Meier, mijn co-promotor, en Betty Coussens, beiden van DSM Research wil ik bedanken voor de inspirerende discussies over dit werk. Prof. Gert Vriend wil ik vooral bedanken voor zijn inspanningen achter de schermen, zonder welke dit proefschrift het licht wellicht niet had gezien. Mijn kamergenoot en CMS-c collega Paul Verwer bedank ik voor zijn immer eloquente en taalkundig opvoedende gesprekken doorspekt met zijn zeer eigen humor. Hens Borkent wil ik bedanken voor zijn goede suggesties met betrekking tot het Molden pakket. Daarnaast wil ik natuurlijk de rest van mijn collega's bij het CMBI bedanken. Ondanks woelige tijden is de goede band is gebleven. De collega's binnen het CMS-c project wil ik bedanken voor een goede samenwerking.

Gijs

Curriculum Vitae

Gijsbert Schaftenaar werd op 28 augustus 1956 in Harderwijk geboren. In juni 1975 behaalde hij het diploma Atheneum-B aan het Christelijk College Nassau Veluwe te Harderwijk. Aansluitend begon hij met de studie scheikunde aan de Rijksuniversiteit Utrecht. Het doctoraalexamen omvatte analytische chemie (Prof. Dr. J. van der Maas) als hoofdvak en kwantumchemie als bijvak. Op 8 maart 1982 werd het doctoraalexamen succesvol afgelegd. Een intermezzo van 14 maanden vormde de vervulling van de militaire dienstplicht. Op 28 augustus 1989 voltooide hij, als assistent in opleiding, een tweejarige onderzoekersopleiding op het gebied van de kwantumchemie van kationen (radicalen en waterstofgebrugde complexen) in de gasfase (zie hoofdstuk 5 van dit proefschrift). Sinds 1 augustus 1989 is hij werkzaam bij het Centre for Molecular and Biomolecular Informatics (het CMBI, voorheen CAOS/CAMM Center), aan de Katholieke Universiteit Nijmegen, de eerste twee jaar binnen een samenwerkingsproject van het CAOS/CAMM Center en de werkgemeenschap Quantumtheoretische Chemie. Het doel hiervan was het ontsluiten van kwantumchemische programmatuur voor chemici in het algemeen. In deze periode werd begonnen met het schrijven van het Molden software pakket (Hoofdstuk 2 van dit proefschrift). De daaropvolgende vier jaar werd onder leiding van Dr. J.A.M. Leunissen gewerkt aan de ondersteuning van de bioinformatica service van het Center. Dit hield in het up to date houden van sequentiedatabanken en het programmeren van gebruikerinterfaces. Daarnaast was hij actief als penningmeester en interim-voorzitter van de Research and Development committee van het EMBnet (European Molecular Biology network). Van 1 augustus 1997 tot 1 augustus 2001 was hij werkzaam als projectmedewerker in dienst van NWO-Chemische Wetenschappen, in het kader van het Kristallisatie project uit het Prioriteiten Programma Materialenonderzoek - Computational Material Science, eerst onder leiding van Dr. J.H. Noordik en later onder leiding van Prof. dr. ir. A. van der Avoird en Prof. dr. E. Vlieg.